Universiteit Utrecht

Faculteit Bètawetenschappen

Simplicial Complexes and Persistent Homology

Bachelor Thesis

Robin van de Griend

Wiskunde

Supervisor:

Dr. Álvaro del Pino Gómez

Mathematisch Instituut

12/06/2020

Abstract

Topological data analysis is a novel field of mathematics in which topology is applied to problems in data analysis. In this paper we will study a particular tool from topological data analysis called *peristent homology.*

First, we develop the theory of simplicial complexes and their topological counterparts polyhedra. We then define a particular type of homology theory called *simplicial homology* which, intuitively, captures n-dimensional 'holes' of polyhedra.

As a motivating example for the definition of persistence homology, we will consider the Čech complex of a point cloud $X \subseteq \mathbb{R}^N$. The Čech complex is constructed as a *parametrized* simplicial complex, that for each choice of a parameter gives a simplicial homology.

By looking at how this simplicial homology changes as we go from one choice of the parameter to another, we then define the *persistence homology* of a filtration. The persistence diagram of a filtration turns out to particularly useful representation of persistence homology. Furthermore the space of all persistence diagrams admits a metric called the *bottleneck distance* by which we can compare the persistence diagrams of different filtrations.

Lastly, we show an important result called The Bottleneck Stability which ensures that the bottleneck distance is stable with respect to small perturbations of the filtration.

Contents

Chapter 1

Simplicial Complexes and Polyhedra

1.1 Simplicial Complexes

The definition of a topological space is very general and allows for spaces which might be too "wild" to be of direct interest. A general idea in topology is to define a certain *structure* on topological spaces, such as manifolds or CW-complexes, that restrict the class of spaces to something more manageable.

In the following discussion, we define simplicial complexes as abstract combinatorial objects that will form the category **Csim** of *simplicial complexes*. We then relate this abstract definition to topological spaces, by defining its *geometric realization*. Those topological spaces that are homeomorphic to a geometric realization will form the category **P** of polyhedra which will be the central object of study. In this section, we follow the exposition in [4, Chapters 2 & 3].

We begin our discussion by defining the objects of **Csim**.

Definition 1.1.1. A **simplicial complex** is pair (K, Φ) given by a *finite*[1] set Φ called the *vertex set* and a set K of subsets in Φ, such that:

- For all $x \in \Phi$ we have that $\{x\} \in K$
- If $\sigma \in K$, then for all non-empty subsets $\tau \subseteq \sigma$ we have that $\tau \in K$.

The elements in Φ are called the *vertices* of K, and the elements $\sigma \in K$ are called the *simplices* of K. The non-empty subsets of a simplex σ are called the *faces* of σ. If a simplex contains $n - 1$ elements, we say that it is an n-simplex. A subset $L \subseteq K$ is called a **subcomplex** when L is also a simplicial complex.

[1] It is possible to define infinite simplicial complexes, but in applications, we only consider finite data sets. As such, this definition will be general enough.

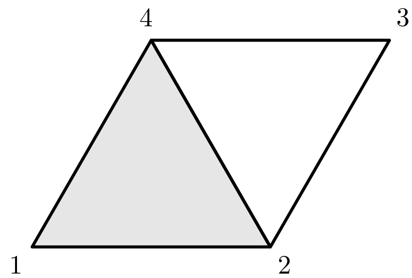

Figure 1.1: A pictorial representation of a simplicial complex.

Similar to how notate topological spaces, we will write K for the simplicial complex (K, Φ) whenever we do not need to specify any particular vertex set.

Example 1.1.2. An example of a simplicial complex is

$$K = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 1\}, \{1, 2, 4\}\}$$

whose vertex set is just the set of integers $\{1, 2, 3, 4\}$. We can pictorially represent this simplicial complex as in figure 2.2. $\triangle$

We now define the morphisms in **Csim**.

Definition 1.1.3. A **simplicial map** on simplicial complexes $(K, \Phi), (L, \Psi)$ is a function $f : K \to L$ such that for all $\sigma = \{x_0, x_1, ..., x_n\} \in K$ we have that $f(\sigma) = \{f(x_0), ..., f(x_n)\} \in L$. That is, f is determined by how it maps the vertices Φ to Ψ and is such that it maps simplices of K to simplices of L.

During the rest of this section, we will build some more theory around simplicial complexes, and in particular we will define its *geometric realization*.

Definition 1.1.4. Two abstract simplicial complexes (K, Φ) and (L, Ψ) are said to be **isomorphic** if there exists a bijective simplicial map $f : K \to L$.

Given a simplicial complex K we can define a new simplicial complex whose vertex set is given by K. That is, the simplices of K will now form the *vertices* of another simplicial complex which we call the barycentric subdivision of K.

Definition 1.1.5. The **barycentric subdivision** of a simplicial complex K is the simplicial complex $\mathcal{B}(K)$ defined by taking K as its vertex set and defining $\mathcal{B}(K)$ as the collection of nonempty subsets $S = \{\sigma_{i_0}, ..., \sigma_{i_n}\}$ of K such that:

$$S \in \mathcal{B}(K) \Leftrightarrow \sigma_{i_0} \subseteq ... \subseteq \sigma_{i_n}$$

This construction turns out to be useful later on when we prove the Nerve Theorem 3.1.4.

Example 1.1.6. Just as with any other simplicial complex, we can pictorially represent a barycentric subdivision $\mathcal{B}(K)$ of K. Doing this with K the simplicial complex from example 1.1.2 gives figure 1.2. We see that the *simplices* of K have now become the *vertices* of $\mathcal{B}(K)$. $\triangle$

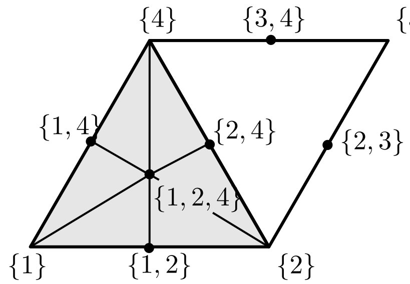

Figure 1.2: A pictorial representation of the barycentric subdivision $\mathcal{B}(K)$ of a simplicial complex K. Each *vertex* of $\mathcal{B}(K)$ is a *simplex* of K.

To describe local behaviour in a simplicial complex K, it might be useful to consider all simplices in K that contain a certain vertex. These simplices form a set, but not nescessarily a simplicial complex, called a *star*. Since simplices are just sets of vertices, we might as well look at the star of a whole simplex, by looking at those simplices that contain its vertices.

Definition 1.1.7. The **star** of a simplex τ in a simplicial complex K is defined as the set $\mathrm{St}(\tau) = \{\, \sigma \in K \mid \tau \subseteq \sigma \,\}$. The **closed star** $\overline{\mathrm{St}}(\tau)$ is the smallest subcomplex of K that contains $\mathrm{St}(\tau)$.

When we collect all the faces τ of a simplex σ we get a new simplicial complex called the simplicial complex generated by σ.

Definition 1.1.8. The **simplicial complex generated by a simplex** $\sigma \in K$ is defined as:

$$\overline{\sigma} := \sigma \cup \mathcal{P}(\sigma) \setminus \emptyset$$

We see that $\overline{\sigma}$ forms the smallest simplicial complex that contains σ. We relate this definition to the closed star of a simplex by the following result:

Lemma 1.1.9. Let $\mathrm{St}(\tau)$ be the star of some simplex τ. Then:

$$\overline{\mathrm{St}}(\tau) = \bigcup_{\sigma \in \mathrm{St}(\tau)} \overline{\sigma}$$

Proof. The inclusion $\overline{\star}(\tau) \subseteq \bigcup_{\sigma \in \mathrm{St}(\tau)} \overline{\sigma}$ is clear. Thus we only need to prove that $\bigcup_{\sigma \in \mathrm{St}(\tau)} \overline{\sigma} \subseteq \overline{\mathrm{St}}(\tau)$. Any element $\rho \in \bigcup_{\sigma \in \mathrm{St}(\tau)} \overline{\sigma}$ is the face of some $\sigma \in \mathrm{St}(\tau)$. Since simplicial complexes are closed under taking faces, $\sigma \in \overline{\mathrm{St}}(\tau)$. □

We will also need the product of two simplicial complexes. Since the vertex sets of simplicial complexes are finite, the vertices always admit an ordering.

Definition 1.1.10. Let (K, Φ) and (L, Ψ) be simplicial complexes, and let Ψ and Φ admit a total order. We define $\Psi \times \Phi$ with the lexicographic ordering induced by Ψ and Φ. Using this ordering, $K \times L$ is defined to be the simplicial complex whose vertices are $\Phi \times \Psi$ and whose simplices are the sets:

$$\{(x_0, y_0), ..., (x_n, y_n)\}$$

such that $\{x_0, ..., x_n\} \in K$ and $\{y_0, ..., y_n\} \in L$.

1.2 Geometric Realization

As it stands, simplicial complexes are built up from simplices and in turn, simplices are defined as combinations of specific vertices. There is a natural geometric interpretation of a simplicial complex, where we envision the simplices as points, lines, triangles, tetrahedrons, etc... and the simplicial complex as a geometric object constructed by glueing these simplices together in a specific manner.

Each point in this geometric object, is contained in a simplex, say for example a triangle with vertices x_1, x_2, x_3. We can now describe this point as a linear combination $\sum \lambda_i x_i$ of these vertices, where the coefficients λ_i add up to 1. This will motivate the following definition, in which we make this notion of a "geometric interpretation" more precise.

Definition 1.2.1. Let (K, Φ) be a simplicial complex. Writing $I := [0, 1]$ we denote by I^Φ the set of all functions $p : \Phi \to I$ mapping the vertices of K to real numbers in I. For each such function, we define its *support* as the finite set:

$$\sup(p) := \{ \sigma \in \Phi \mid p(\sigma) > 0 \}$$

Then, the **geometric realization of a simplicial complex** K is the set:

$$|K| = \{ p \in I^\Phi \mid \sup(p) \in K \text{ and } \sum_{x \in \Phi} p(x) = 1 \}$$

The connection with the preceding discussion is the following. For each vertex $x \in \Phi$ we define a function $x : \Phi \to I$ by:

$$x(y) = \begin{cases} 1 & \text{if } \sigma = x \\ 0 & \text{if } \sigma \neq y \end{cases} \tag{1.2.1}$$

Each $p \in |K|$ with support $\sup(p) = \{x_0, ..., x_n\} \subseteq \Phi$, can now be written as a linear combination of the functions $x_i \in |K|$:

$$p = \sum_i \lambda_i x_i$$

where $\lambda_i := p(\{x_i\})$. And indeed, by definition of the geometric realization: $\sum_i \lambda_i = 1$.

Definition 1.2.2. Let $|K|$ be a geometric realization of a simplicial complex K. For each element $p = \sum_i \lambda_i x_i \in |K|$, we call the coefficients λ_i the **barycentric coordinates** of p.

We now define a metric on $|K|$.

Lemma 1.2.3. The function $d : |K| \times |K| \to \mathbb{R}$ defined as:

$$d(p, q) := \sqrt{\sum_{x \in \Phi} (p(x) - q(x))^2}$$

is a metric on the geometric realization $|K|$ of a simplicial complex K.

Proof. Since Φ is finite, we see that $d(p, q)$ is just the euclidean norm on the vector $(p(x_0) - q(x_0), ..., p(x_n) - q(x_n))$. □

We see now that each simplicial complex $K \in \mathbf{Csim}$, maps to a topological space $|K| \in \mathbf{Top}$. Similarly, we will now define a way to map each simplicial map $f : K \to L$ to a continuous map $|f| : |K| \to |L|$ so that we get a *functor* $|\cdot| : \mathbf{Csim} \to \mathbf{Top}$.

Definition 1.2.4. The **geometric realization of a simplicial map** $f : K \to L$ is the function $|f| : |K| \to |L|$ defined by:

$$|f|(\sum_i \lambda_i x_i) = \sum_i \lambda_i f(x_i)$$

By the following theorem, which will be stated without proof but which can be found in [4, Thm.2.2.7], $|f|$ is indeed continuous so that $|\cdot|$ defines a functor $\mathbf{Csim} \to \mathbf{Top}$.

Theorem 1.2.5. Given a simplicial map $f : K \to L$, its geometric realization $|f|$ is continuous.

Another result that we will not prove here is the following (see [4, Thm. 3.1.1]).

Theorem 1.2.6. Given two simplicial complexes K and L the product of their geometric realizations is homeomorphic to the geometric realization of their product:

$$|K \times L| \simeq |K| \times |L|$$

We can also consider the geometric realization of the barycentric subdivision of a simplicial complex. As we saw in example 1.1.6, the pictorial representation of $\mathcal{B}(K)$ topologically looks the same as the one for K and thus the following result will not be surprising. For a proof we refer the reader to [4, Thm. 2.1.4].

Lemma 1.2.7. Given a simplicial complex K and its barycentric subdivision $\mathcal{B}(K)$, their geometric realizations are homeomorphic:

$$|K| \simeq |\mathcal{B}(K)|$$

As stated at the beginning of this section, we defined a *structure*, namely simplicial complexes, by which we can restrict ourselves to a specific class of topological spaces called *polyhedra*.

Definition 1.2.8. A **polyhedron** is a topological space X that is homeomorphic to the geometric realization $|K|$ of some simplicial complex K. We denote by $\mathbf{P}$ the category of polyhedra with continuous functions as the morphisms.

We should note that not every continuous map between polyhedra is the geometric realization of some simplicial map, so there are many more morphisms in $\mathbf{P}$ than in $\mathbf{Csim}$.

In the next section, we will define a topological invariant on polyhedra called *homology*, which we will use to captures the “holes” in polyhedra. As it turns out, it is enough to just look at the simplicial structure underlying the polyhedron, such that we will just define homology for simplicial complexes.

Chapter 2

Simplicial Homology

2.1 Capturing 'Holes'

The idea of simplicial homology, is to define an algebraic object that in some sense captures n-dimensional holes in a polyhedron $|K|$. More precisely, for each dimension, we will define a group, generated by the simplices in K, that describes each hole as an element of the group.

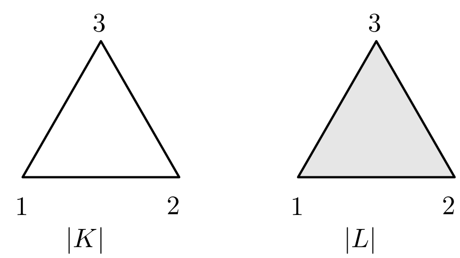

Figure 2.1: The geometric realizations of K and L

Take for example the simplicial complex $K = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{3, 1\}\}$ with vertex set $\{1, 2, 3\}$. Its geometric realization can be seen in figure 2.1. As a simplicial complex it consists simply of the three 1-simplices in the boundary of the 2-simplex $\{1, 2, 3\}$ and the 0-simplices $\{1\}, \{2\}, \{3\}$. By following the triangle along the edges we end up in the same point, so if we take the ordering of vertices in consideration the product $\{1, 2\}\{2, 3\}\{3, 1\}$ in some sense describes a *loop* around the hole in K, where the ordering of the vertices indicates in which way we follow the edge.

But what about the products $\{2, 3\}\{3, 1\}\{1, 2\}$ and $\{3, 1\}\{1, 2\}\{2, 3\}$? Like $\{1, 2\}\{2, 3\}\{3, 1\}$ they describe loops around the 'hole' of K, but their start/end-point is different. These three products describe the same unparametrized circle around K, so if we are trying to capture 'holes', it would be natural to consider these three products to be the same:

$$\{1, 2\}\{2, 3\}\{3, 1\} = \{2, 3\}\{3, 1\}\{1, 2\} = \{3, 1\}\{1, 2\}\{2, 3\} \tag{2.1.1}$$

But what we see now is that each edge commutes with the other two. This suggest that the group we want to define should actually be Abelian, and that we should talk of 'sums' instead of 'products'. What we have done here, is to forget any specific parametrization of the loops to get the unparametrized circles, also called *cycles*, that actually describe *holes* in X. For those who are familiar with the fundamental group of a topological space, we will see that this group is the Abelianization of the first fundamental group.

Now lets consider $L = K \cup \{1,2,3\}$, such that the only difference with $K \subseteq L$ is that the hole surrounded by the edges in L has now become 'filled in' by the 2-simplex $\{1,2,3\}$ and disappeared. To algebraically describe this disappearance, we somehow want to 'quotient' all the cycles of 1-simplices in L by the 2-simplices which they bound.

2.2 Chain Groups

To make this all precise, we will first define what it means for a simplicial complex to have an *orientation*.

Definition 2.2.1. Given an n-simplex $\sigma = \{x_0, ..., x_n\}$ in (K, Φ), we can order its vertices $x_0, ..., x_n$ in $(n+1)!$ different ways. An **orientation of a simplex** σ is an equivalence class of orderings of its vertices such that:

$$x_0, ..., x_n \sim p(x_0), ..., p(x_n) \Leftrightarrow p \text{ is an even permutation}$$

Thus, we say that two orderings have the same orientation when they differ by an even permutation. Then we call σ together with an orientation an **oriented simplex**. For $n > 0$, each n-simplex has two different orderings so that we can write $\sigma = -\tau$ when σ and τ are the same as simplices, but have opposing orientation.

For example, we can write $\{1,2,3\} = \{2,3,1\} = -\{3,2,1\}$. We now define what an orientation is on a whole simplicial complex:

Definition 2.2.2. An **oriented simplicial complex** is a simplicial complex whose simplices all have an orientation.

Any total ordering of the vertices Φ of a simplicial complex K, induces an orientation on each simplex $\sigma \in K$. Since we only consider *finite* Φ, we can make always choose an orientation on K.

Definition 2.2.3. For each $n \in \mathbb{N}$ we define the **n-th chain group** $C_n(K)$ of a simplicial complex K as the free Abelian group generated by the n-simplices in K. The elements of $C_n(K)$ are called **n-chains** and can be written as formal sums $\sum_i n_i \sigma_i$ with each σ_i an n-simplex of K and $n_i \in \mathbb{Z}$.

The n-chains are the actual objects that we will use to represent holes in a simplicial complex. Later on we will define a slight generalization of simplicial homology, where the coefficients n_i lie in any commutative ring.

Example 2.2.4. Recall the simplicial complexes $K = \{\{1\},\{2\},\{3\},\{1,2\},\{2,3\},\{3,1\}\}$ and $L = K \cup \{1,2,3\}$ that we considered in figure 2.1. We remarked that the simplices $\{1,2\},\{2,3\},\{3,1\}$ formed the boundary of $\{1,2,3\}$ in L, corresponded to a 'hole' in some way in K. In light of this, we can define the boundary of $\{1,2,3\}$ as the 1-chain $\{1,2\}+\{2,3\}+\{3,1\}$ in L. △

This motivates the following definition of the *boundary homomorphism*. We will define it on chain groups by giving its value on the generators:

Definition 2.2.5. Given two chain groups $C_n(K), C_{n-1}(K)$, we define the n-th **boundary homomorphism** as the map

$$\partial_n : C_n(X) \to C_{n-1}(X) : \{v_0, ..., v_n\} \mapsto \sum_i (-1)^i \{v_0, ..., \hat{v}_i, ..., v_n\} \tag{2.2.1}$$

for $n > 0$ and as $\partial_0 : C_0(X) \to 0$ for $n = 0$ and where $\{v_0, ..., \hat{v}_i, ..., v_n\}$ is the simplex $\{v_0, ..., v_n\}$ where we removed the vertex v_i.

The boundary homomorphism sends an n-simplex σ to a linear combination of its faces that represents the boundary. The orientation that we give to simplices in the resulting n-chain $\partial(\sigma)$ is a natural orientation induced from the orientation on σ.

Example 2.2.6. Recall the simplicial complexes $K = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{3, 1\}\}$ and $L = K \cup \{1, 2, 3\}$ that we considered in figure 2.1. Now we have defined the boundary of $\{1, 2, 3\}$ in L as the 1-chain $\{1, 2\} - \{1, 3\} + \{2, 3\} = \{1, 2\} + \{2, 3\} + \{3, 1\}$, where we see that the induced ordering on the simplices in the chain correspond to our intuition of how the edges of $\{1, 2, 3\}$ are ordered. △

Heuristically, the border of a simplex $\partial\sigma$ is always a closed cycle around the interior of σ and as such, the border of a border is empty. To make this statement precise, we prove the following lemma:

Lemma 2.2.7. The composition $\partial_{n-1} \circ \partial_n$ is the zero map.

Proof. Let $\sigma = \{v_0, ..., v_n\}$ an n-simplex. Notice that for $j > i$, the $(j-1)$-th vertex in $\{v_0, ..., \hat{v}_i, ..., v_n\}$ is v_j, so that:

$$\begin{aligned}\partial_{n-1} \circ \partial_n(\sigma) &= \sum_i (-1)^i \partial_{n-1} \{v_0, ..., \hat{v}_i, ..., v_n\} \\ &= \sum_{j<i} (-1)^i (-1)^j \{v_0, ..., \hat{v}_j, ..., \hat{v}_i, ..., v_n\} + \sum_{j>i} (-1)^i (-1)^{j-1} \{v_0, ..., \hat{v}_i, ... \hat{v}_j, ..., v_n\}\end{aligned}$$

If we swap i and j in the second summation, it cancels out with the first summation to give the desired result. □

We get a sequence of chain groups, called a **chain complex**, connected by the corresponding boundary homomorphisms in the following way:

$$... \xrightarrow{\partial} C_2 \xrightarrow{\partial} C_1 \xrightarrow{\partial} C_0 \xrightarrow{0} 0 \tag{2.2.2}$$

Notice how we have omitted the indexing on the boundary homomorphisms. We will do this when it is clear from context which map we mean.

It would be natural to look at maps between chain complexes of two simplicial complexes K and L, induced by some simplicial map between K and L themselves.

Definition 2.2.8. Let K and L be two simplicial complexes, and $f : K \to L$ a simplicial map. Then we define the **induced homomorphism** of f as the map:

$$f_\# : C_n(K) \to C_n(L) : \sum_i n_i\sigma_i \mapsto \sum_i n_i f(\sigma_i)$$

Using this definition we have the following result.

Lemma 2.2.9. For an induced map $f_\# : C_n(K) \to C_n(L)$, the following diagram commutes:

$$\begin{array}{ccc} C_n(K), & \xrightarrow{\partial} & C_{n-1}(K) \\ \downarrow{\scriptstyle f_\#} & & \downarrow{\scriptstyle f_\#} \\ C_n(L) & \xrightarrow{\partial} & C_{n-1}(L) \end{array}$$

That is, $\partial f_\# = f_\# \partial$.

Proof. Let $\sigma = \{v_0, ..., v_n\}$ be a generator in $C_n(K)$. Then:

$$\begin{aligned} f_\#\partial(\sigma) &= \sum_i (-1)^i f_\#(\{v_0, ..., \hat{v}_i, ..., v_n\}) \\ &= \sum_i (-1)^i \{f(v_0), ..., f(\hat{v_i}), ..., f(v_n)\}) \\ &= \partial(f(\sigma)) = \partial f_\#(\sigma) \end{aligned}$$

Where the first equation follows from the linearity of $f_\#$ in the generators of C_n. □

2.3 Simplicial Homology Groups

For a simplicial complex K, we will now define its **n-th homology group** $H_n(K)$ in terms of quotient classes of n-chains. If a cycle is not surrounding a 'hole', but is actually the boundary of a larger simplex, we would like to disregard it inside $H_n(K)$. We have that $\mathrm{Im}(\partial_n) \subseteq \ker(\partial_{n-1})$ by 2.2.7. Since the image and kernel of a group homomorphism are groups and subgroups of Abelian groups are normal, the quotient of these two groups is well-defined so that the following definition makes sense:

Definition 2.3.1. The n-th **homology group** of a simplicial complex K is defined as:

$$H_n(K) = \ker(\partial_n)/\,\mathrm{Im}(\partial_{n+1})$$

and we denote its *rank* by β_i which we will call the n-th Betti number of K.

We can easily extend the definition to *polyhedra* as follows:

Definition 2.3.2. Let $|K|$ be a polyhedron, homeomorphic to the geometric realization of a simplicial complex K. Then its n-th homology group is defined as:

$$H_n(|K|) := H_n(K)$$

Similar to what we did for Chain groups, we show that each simplicial map $f : K \to L$ induces a group homomorphism on the homology groups of K and L.

Lemma 2.3.3. Every simplicial map $f : K \to L$ induces a homomorphism

$$f_* : H_n(K) \to H_n(L) : [\sigma] \mapsto [f_\#(\sigma)]$$

Proof. We will show that f_* is well-defined. Let $[\sigma] = [\tau]$ in $H_n(K)$, so $\sigma, \tau \in \ker \partial_n$ and $\sigma - \tau \in \operatorname{Im} \partial_{n+1}$. Using 2.2.9, $\partial_n f_\#(\sigma) = f_\# \partial_n(\sigma) = 0$, so $f_\#(\sigma) \in \ker \partial_n \subseteq C_n(L)$. Similarly: $f_\#(\tau) \in \ker \partial_n$. Since $\sigma - \tau \in \operatorname{Im} \partial_n$ there exists a $\mu \in C_n(K)$ such that $\sigma - \tau = \partial_n(\mu)$. Thus: $\partial_n f_\#(\mu) = f_\# \partial_n(\mu) = f_\#(\sigma - \tau) \in C_n(L)$ so $f_\#(\sigma - \tau) = f_\#(\sigma) - f_\#(\tau) \in \operatorname{Im} \partial_n$. We conclude that $f_*([\sigma]) = f_*([\tau])$ in $H_n(L)$. That f_* is a homomorphism follows directly from the fact that $f_\#$ is a homomorphism. □

Some basic properties of the induced homomorphism are the following:

- $(fg)_* = f_* g_*$, which follows from the associativity of function composition.
- $\mathrm{Id}_* = \mathrm{Id}$, which follows from: $\mathrm{Id}_*[\sigma] = [\mathrm{Id}_\#(\sigma)] = [\sigma]$.

It would now be natural to ask if any map $|K| \to |L|$ between polyhedra also induces a homomorphism $H(|K|) \to H(|L|)$. Surely, any map $|f| : |K| \to |L|$ that is a geometric realization of a simplicial map $f : K \to L$ induces the homomorphism $f_* : H(K) \to H(L)$. But since not every map between polyhedra is the geometric realization of some simplicial map between simplicial complexes, we can not easily extend this definition.

Luckily, for each continous map $f : |K| \to |L|$ between polyhedra we can find a simplicial map $g : \mathcal{B}^r(K) \to L$ from the r-fold barycentric subdivision $\mathcal{B}^r(K)$ to L that 'approximates' the continous map 'well enough'. By well enough we mean that the induced map g_* composed with a certain isomorphism between $H(\mathcal{B}^r(K))$ and $H(K)$ defines a unique isomorphism $f_* : H(K) \to H(L)$. We summarize in the following theorem, which will be stated without proof (see [4, Thm. 3.2.7]).

Theorem 2.3.4. Any continuous map $f : |K| \to |L|$ induces a unique homomorphism $f_* : H(|K|) \to H(|L|)$. Given another continuous map $g : |K| \to |L|$, the identity $(fg)_* = f_* g_*$ holds.

2.4 Homotopy Invariance

Before we continue with proving some more things about simplicial homology, we will first consider the larger context in which we are working. We have stated the homology theory in terms of simplicial complexes and have thereby restricted ourselves to polyhedra. There are ways to make the definition more general, so that we can consider arbitrary topological spaces, but as we will see, the definition we have given will be general enough. The idea is that, if we hope to find any interesting structures in our data, and if we hope this structure to contain any sort of qualitative information, it should at least take on a *simplicial* structure. The more general case allows for topological spaces that are, as stated before, too "wild" to be of interest.

Even so, the dataset X we are working with, might not directly be given as a simplicial complex. In chapter 3, we will describe a process by which we can obtain a simplicial complex from X. Before this makes sense though, we should ask ourselves if any simplicial structure of X sufficiently "captures" the homology of the underlying structure of the data. Mathematically speaking, we're asking whether two homotopy equivalent polyhedra also have the same homology groups. Indeed, this is the case and we will prove this in the following section. Proof will be based on the exposition in [6, Chapter 2.1] and [4, Thm. 2.3.0].

Theorem 2.4.1. If $f, g : |K| \to |L|$ are homotopic maps on simplicial complexes $|K|$ and $|L|$, then they induce the same maps $g_*, f_* : H_n(|K|) \to H_n(|L|)$.

Proof. Assume we have a homotopy $H : |K| \times I \to |K|$ such that $H(p, 0) = f(p)$ and $H(p, 1) = g(p)$. We can take the unit interval $I = [0, 1]$ as the geometric realization of a simplicial complex $\mathcal{I} := \{\{0\}, \{1\}, \{0, 1\}\}$. Now, we define two simplicial maps:

$$i_0 : K \to K \times \mathcal{I} : x \mapsto (x, 0)$$
$$i_1 : K \to K \times \mathcal{I} : x \mapsto (x, 1)$$

Then under the identification $|K \times \mathcal{I}| \simeq |K| \times I$ given by theorem 1.2.6 the geometric realizations of these maps are given for i_0 by:

$$|i_0| : |K| \to |K \times \mathcal{I}| : p = \sum_i \lambda_i x_i \mapsto \sum_i \lambda_i(x_i, 0) = (\sum_i \lambda_i x_i, 0) = (p, 0)$$

and similarly by $|i_1| : p \mapsto (p, 1)$ for i_1. By composing the map i_0 with H we see that

$$f_* = (H \circ |i_0|)_* = H_*|i_0|_*$$
$$g_* = (H \circ |i_1|)_* = H_*|i_1|_*$$

Thus, to prove the theorem we only need to show that $(i_0)_* = |i_0|_*$ and $|i_1|_* = (i_1)_*$ are equivalent maps from $H_n(K)$ to $H_n(K \times \mathcal{I})$. To do this we will prove that for all $\alpha \in C_n(K)$, $(i_0)_*(\alpha) - (i_1)_*(\alpha)$ is the boundary of some chain in $C_{n+1}(K \times \mathcal{I})$. To that end we will define a map $P : C_n(K) \to C_{n+1}(K \times (I)$ that takes an n-simplex $\sigma \in K$ and maps it to a linear combinations of simplices in $K \times \mathcal{I}$.

Given an ordering of the simplices $x_0, ..., x_r$ of K, we define for $0 \leq r + 1$ simplicial maps

$$f_k : K \to K \times \mathcal{I} : \{x_i\} \mapsto \begin{cases} (x_i, 0) & \text{if } i < k \\ (x_i, 1) & \text{if } i \geq k \end{cases}$$

Note that $f_0 = i_0$ and $f_{r+1} = i_1$. Intuitively, we can envision $K \times \mathcal{I}$ as a *prism*, with K as the top and bottom. By moving the last vertex $(x_n, 0)$ of the n)-simplex $\sigma = \{(x_0, 0), ..., (x_n, 0)\} \in K \times 0$ up to $(x_n, 1)$, the space between the bottom and top simplex now form a new $(n + 1)$-simplex $f_n(\sigma) = \{(x_0, 0), ..., (x_n, 0), (x_n, 1)\}$. If we do the same to the n-simplex $\{(x_0, 0), ..., (x_n, 1)\}$ we obtain the simplex $f_{n-1}(\sigma)$. As we continue this process, we interpolate between $f_0 = i_0$ and $f_{r+1} = i_1$ through simplicial maps f_n, similar to how a homotopy between two continuous maps interpolates by continuous maps.

We will now define the *prism operator* $P(\sigma) : C_n(K) \to C_{n+1}(K \times \mathcal{I})$ that we hinted at in the beginning of the proof:

$$P(\sigma) = \sum_k (-1)^k f_k(\sigma) \tag{2.4.1}$$

Heuristically, the top edge and bottom edge of P will correspond to $(i_0)_\#$ and $(i_1)_\#$ respectively. This motivates the following identity:

$$\partial P = (i_0)_\# - (i_1)_\# - P\partial \tag{2.4.2}$$

Before we derive this identity, we'll show that it proves the theorem. Let $[\alpha] \in H_n(K)$, so in particular $\alpha \in \ker \partial_n$. Then using 2.4.2:

$$\partial P(\alpha) = (i_0)_\#(\alpha) - (i_1)_\#(\alpha)$$

so $(i_0)_\#(\alpha) - (i_1)_\#(\alpha) \in \operatorname{Im} \partial_{n+1}$ meaning they are equivalent in $H_n(K)$, proving the theorem.

The last step is to prove, rather technically, identity 2.4.2. By striking through those vertices that are removed from the simplex by the boundary map, we calculate:

$$\begin{aligned}\partial P(\sigma) = &\sum_{l\le k}(-1)^l(-1)^k\{(x_0,0),...,\cancel{(x_l,0)},...,(x_k,0),(x_k,1),...,(x_n,1)\}\\ &+\sum_{l\ge k}(-1)^l(-1)^k\{(x_0,0),,...,(x_k,0),(x_k,1),...,\cancel{(x_l,1)}...,(x_n,1)\}\end{aligned}$$

and

$$\begin{aligned}P\partial(\sigma) = &\sum_{l<k}(-1)^{l-1}(-1)^k\{(x_0,0),...,\cancel{(x_l,0)},...,(x_k,0),(x_k,1),...,(x_n,1)\}\\ &+\sum_{l>k}(-1)^l(-1)^k\{(x_0,0),,...,(x_k,0),(x_k,1),...,\cancel{(x_l,1)}...,(x_n,1)\}\end{aligned}$$

so that:

$$\partial P(\sigma) - P\partial(\sigma) = \sum_l (-1)^{2l}\{(x_0,0),,...,\cancel{(x_l,0)},(x_l,1),...,(x_n,1)\} \tag{2.4.3}$$

$$- \sum_l (-1)^{2l}\{(x_0,0),,...,(x_l,0),\cancel{(x_l,1)},...,(x_n,1)\} \tag{2.4.4}$$

But since $\{(x_0,0),,...,\cancel{(x_l,0)},(x_l,1),...,(x_n,1)\} = \{(x_0,0),,...,(x_l,0),\cancel{(x_l,1)},...,(x_n,1)\}$ for all l except 0 and n, the terms in 2.4.3 cancel out to give:

$$\partial P(\sigma) - P\partial(\sigma) = f_n(\sigma) - f_0(\sigma) = (i_0)_\#(\sigma) - (i_1)_\#(\sigma)$$

as desired. □

We conclude with the following important corollary.

Corollary 2.4.2. If $|K|$ and $|L|$ are homotopy equivalent by a map $f : |K| \to |L|$, then the homomorphism $f_* : H_n(|K|) \to H_n(|L|)$ induced by f is an isomorphism for all n.

Proof. Since f is a homotopy equivalence, there exists a map $g : |L| \to |K|$ such that $gf \simeq \mathrm{Id}_{|K|}$ and $fg \simeq \mathrm{Id}_{|L|}$. By 2.4.1 and the basic properties of the induced homomorphism stated above: $g_*f_* = (gf)_* = (\mathrm{Id}_{|K|})_* = \mathrm{Id}$. Similarly we prove that $f_*g_* = \mathrm{Id}$. So f_* is an invertible homomorphism, hence an isomorphism. □

This result states what we hoped: that homology is a homotopy invariant property of a polyhedron. It gives us the theoretic guarantee that as long as our process of obtaining simplicial complexes from a dataset is determined up to homotopy type, it will also be determined up to homology.

2.5 Homology with Coefficients

As you will recall, n-chains, the elements of the chain group $C_n(K)$, are defined as linear combinations $\sum n_i\sigma_i$ with *coefficients* n_i in $\mathbb{Z}$. Heuristically, the sign of the coefficients describe the orientation of simplices within the n-chain and the magnitude the number of times this simplex is represented in the n-chain.

We define a generalization of the chain group by allowing the coefficients to lie in any Abelian group G instead of just $\mathbb{Z}$, and we will write the resulting chain group as $C_n(K, G)$. In fact, since in the discussion above we have only used the fact that $\mathbb{Z}$ is Abelian, all the theory remains valid. In particular, the definition of the boundary homomorphism ∂ remains valid so that the resulting homology groups $H_n(K, G)$ remain well-defined.

One particular case we will be using in the calculation of persistent homology is $G = \mathbb{Z}_2$. Let us first consider how the boundary homomorphism looks like on $\alpha = (\sum_i n_i\sigma_i) \in C_n(K, \mathbb{Z}_2)$:

$$\partial(\alpha) = \sum_i n_i\partial(\sigma_i) = \sum_{i,j}(-1)^j n_i\{x_0, ..., \hat{x}_j, ..., x_n\}$$

But since $\partial(\alpha) \in C_{n-1}(K, \mathbb{Z}_2)$, its coefficients $(-1)^j n_i$ must also lie in $\mathbb{Z}_2 = \{0, 1\}$. This means we can simplify the whole definition to give:

$$\partial(\alpha) = \sum_i n_i\partial(\sigma_i) = \sum_{i,j} n_i\{x_0, ..., \hat{x}_j, ..., x_n\} \tag{2.5.1}$$

A second observation to make is that since $\mathbb{Z}_2$ is a *field*, $C_n(K, \mathbb{Z}_2)$ actually becomes a $\mathbb{Z}_2$-*vector space*! Before, the n-simplices of K formed a basis of $C_n(K)$ in a group theoretic sense, now they are still a basis of $C_n(K, \mathbb{Z}_2)$, but in the sense of a vector space.

2.6 Calculating Homology

We will continue with the case of coefficients in $\mathbb{Z}_2$. In this case, the boundary map $\partial_n : C_n \to C_{n-1}$ is now a linear map between vector spaces and can be realized as multiplication with a $d_{n-1} \times d_n$ matrix, where $d_i = \dim C_i$. We choose an ordering of the simplices, which gives us an ordered basis of C_n and C_{n-1}. Since this linear map sends each n-simplex to its boundary $(n-1)$-simplices, its associated matrix becomes $\partial_n = (a_{ij})$ where $a_{ij} = 1$ if the i-th simplex in C_{n-1} is in the boundary of the j-th simplex in C_n, and $a_{ij} = 0$ if this is not the case. For any n-chain $c = \sum a_i\sigma_i$, we calculate its boundary $\partial_n c$ by writing it in vector form $c = (a_1, ..., a_{d_n})$ and multiplying it with the matrix ∂_n.

As an example, let us consider the simplicial complex K shown in 2.2. It consists of four 0-simplices labeled from 1 to 4, five 1-simplices, and a single 2-simplex. If we order the simplices

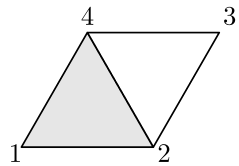

Figure 2.2: A simplicial complex K. Note that the greyed out triangle signifies the 2-simplex $\{1, 2, 4\}$.

according to the lexicographic ordering on the labels of the vertices, we can write down the boundary matrices as:

$$\partial_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \qquad \partial_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

More precisely, ∂_1 is the matrix of the first boundary homomorphism with respect to the basis $\{\{1, 2\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\} \subseteq C_1(K)$ and ∂_2 is the matrix of the 2-nd boundary homomorphism in relation to the basis $\{\{1, 2, 4\}\} \subseteq C_2(K)$.

From this formulation, we can now calculate the Betti numbers of X by considering the dimensions of the kernel and the image of the boundary matrices ∂_n:

$$\beta_n = \dim(\ker \partial_n / \operatorname{Im} \partial_{n+1}) = \dim(\ker \partial_n) - \dim(\operatorname{Im} \partial_{n+1})$$

To algorithmically obtain the dimensions of the kernel and image of a matrix, we would like to reduce the matrix to its *Smith normal form*, that is:

Definition 2.6.1. A matrix $A = (a_{ij}) \in \mathbb{Z}_2^{n \times m}$ is in Smith normal form when:

- For all $i \neq j$, $a_{ij} = 0$, that is: A is a diagonal matrix.
- For some $0 \leq k \leq n$ we have that $a_{ii} = 1$ for $i \leq k$ and $a_{ii} = 0$ for $i \geq k$.

A matrix A in Smith normal form is a diagonal matrix with $\operatorname{Diag}(A) = (1, ..., 1, 0, ..., 0)$. For example:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.6.1}$$

We can now easily read of $\dim(\ker A) = 1$ and $\dim(\operatorname{Im} A) = 2$ as the number of empty columns and non-empty rows respectively.

We will now consider an algorithm for the specific case of matrices with entries in $\mathbb{Z}_2$ which is similar to Gaussian elimination. If $B \in \mathbb{Z}_2^{n \times m}$ is the matrix that we want to reduce to its Smith normal form A, we will do this by applying a chain of elementary row and column operations to B. In terms of matrices, we are looking for invertible matrices $P \in \mathbb{Z}_2^{n \times n}$ and $Q \in \mathbb{Z}_2^{m \times m}$ such that $A = PBQ$. The specific elementary operations we will use are the interchanging of rows/columns and the addition of a row/column to another row/column. From linear algebra we know that elementary matrices are invertible and thus do not change the dimension of the kernel and image of B. In other words, for any matrix of ∂_n, we can extract the Betti number by looking at the matrix in Smith normal form. In the outline of the algorithm, instead of specifying the elementary matrices with which we multiply B, we will just specify which elementary operation we are doing.

Theorem 2.6.2. Let $B = (b_{ij})$ be any $n \times m$ matrix with entries in $\mathbb{Z}_2$. Then B can be reduced to a matrix $A \in \mathbb{Z}_2^{n\times m}$ by elementary row and column operations

Proof. If $B = 0$ it is already in Smith normal form and we are done. If not, there is an element $b_{ij} = 1$. By a consecutively swapping the i-th row with the 1-st row and the j-th column with the 1-st column, we get $b_{11} = 1$. Now for each element $b_{i1} = 1$ in the same *column* as b_{11}, we simple add the 1-st row to the j-th row to get $b_{i1} = 1$. We do this similarly for each $b_{1j} = 1$ in the same *row* as b_{11} to get a matrix with block form:

$$\begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix}$$

For C some $n-1 \times m-1$ matrix in $\mathbb{Z}_2$. We continue this process recursively, by applying the same algorithm to C, to ultimately end up with a matrix in Smith normal form. $\square$

More generally we could have defined the Smith normal form for matrices with entries in any ring. If this ring is also a *principial ideal domain* we can always reduce these matrices to its Smith normal form.

If we apply the above algorithm to our example K from figure 2.2 we get the following matrices for ∂_1 and ∂_2:

$$\partial_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \partial_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

From which we see: $\beta_1 = \dim(\ker \partial_1) - \dim(\operatorname{Im} \partial_2) = 2 - 1 = 1$. This agrees with our intuition in that the first Betti number is equal to the number of "1-dimensional holes" in K.

In the next chapter, we will look at the situation where we don't just have *one* simplicial complex K, but a whole *chain* of complexes $K_0 \subseteq K_1 \subseteq ... \subseteq K_n$. In this case, we want know how the homology classes of the complexes change as we go along this chain.

Chapter 3

Persistent Homology

In the following section, we will turn our discussion towards the study of a certain finite point cloud $S \subseteq \mathbb{R}^n$. We will see how we can study the *shape* of S by using homology. Since S is finite set in $\mathbb{R}$, it does not yet contain any interesting homology. So first we will construct a simplicial complex from S that in some sense *approximates* its shape. More generally, we can construct a simplicial complex from a cover of a topological space X, by taking its *nerve*, which we prove is homotopy equivalent to the cover.

Then, going back to the specific case of our point cloud S, we construct a parametrized cover of S, whose nerve is called the Čech complex of S. For each choice of the parameter, we can calculate the simplicial homology of the Čech complex. By looking at how the simplicial homology changes as we go from one choice of the parameter to another, we define its *persistence homology*. The persistence homology turns out to have a nice representation called the persistence diagram, which we will prove is stable with respect to small perturbations in the point cloud S.

3.1 The Nerve Theorem

Say we have a topological space X and U_0, U_1, U_2 subsets of X. If U_0 intersects with U_1, we might register this fact by writing $\{U_0, U_1\}$. Similarly, if the three sets U_0, U_1, U_2 have non-empty intersection, we could register this by writing $\{U_0, U_1, U_2\}$. By doing this, we are describing the different ways by which subsets of a topological space X intersect in terms of sets. What we will see is that these sets actually form a simplicial complex.

Definition 3.1.1. For a topological space X, we consider a finite collection $\mathcal{U}$ of subsets $U \subseteq X$. The **nerve** $\mathcal{U}$, is the collection of all subsets of $\mathcal{P}(\mathcal{U})$ whose sets have non-empty intersections:

$$\mathcal{N}(\mathcal{U}) := \left\{ X \subseteq \mathcal{P}(\mathcal{U}) \;\middle|\; \bigcap X \neq \emptyset \right\}$$

So, simplices in $\mathcal{N}(\mathcal{U})$ correspond to subcollections of sets in $\mathcal{U}$ that have non-empty intersection. This indeed defines a simplicial complex:

Lemma 3.1.2. For any finite collection of sets $\mathcal{U}$, its nerve $\mathcal{N}(\mathcal{U})$ defines a simplicial complex.

Proof. Let $X \in \mathcal{N}(\mathcal{U})$, then if $Y \subseteq X$, we have that for all $S \in Y$ also $S \in X$. So: $\bigcap X \subseteq \bigcap Y$ and since $\bigcap X$ is non-empty, $\bigcap Y$ is also non-empty. $\square$

So we are able to map a topological space X, together with a collection $\mathcal{U}$ of sets in X to a simplicial complex. If X is embeddable into $\mathbb{R}^N$, and $\mathcal{U}$ is a collection of closed convex sets that covers X, then X turns out to be homotopy equivalent to the geometric realization of its nerve. The proof we present here is based on the original proof by André Weil [9] (for a translation, see [7]).

Given a simplex $\sigma \in K$, its closed star $\overline{\mathrm{St}}(\sigma)$ is a simplicial complex so that we can consider its geometric realization $|\overline{\mathrm{St}}(\sigma)|$. The following lemma relates this realization to points in $|\mathcal{B}(K)|$ and will be needed for proving the Nerve Theorem.

Lemma 3.1.3. Let (K, Φ) be a simplicial complex with vertex set Φ and $B := (\mathcal{B}(K), K)$ its barycentric subdivision. Consider the simplex $\{x\} \in K$ consisting of a single vertex $x \in \Phi$, then:

$$|\overline{\mathrm{St}}(\{\{x\}\})| = \{p \in |B| \mid \{x\} \subseteq \tau \text{ for all } \tau \in \sup p\} \tag{3.1.1}$$

where $\overline{\mathrm{St}}(\{\{x\}\})$ is the closed star of the element $\{\{x\}\} \in B$.[1]

Proof. We will write $\mathrm{St}(x)$ as a shorthand for $\mathrm{St}(\{\{x\}\}) \subseteq B$.

Let $p \in |B|$ such that $\{x\} \subseteq \tau$ for all $\tau \in \sup p$. Thus, by definition, $\sup(p) \subseteq \mathrm{St}(x) \subseteq \overline{\mathrm{St}}(x)$ and thus $p \in |\overline{\mathrm{St}}(x)|$.

Now let $p \in |\overline{\mathrm{St}}(x)|$. By lemma 1.1.9 we have:

$$|\overline{\mathrm{St}}(x)| = \bigcup_{\sigma \in \mathrm{St}(x)} |\overline{\sigma}|$$

so that $p \in |\overline{\sigma}|$ for some $\sigma \in \mathrm{St}(x)$. Since $\{x\} \in \sigma$ and since no other $\tau \in K$ is contained in $\{x\}$, by the definition of the barycentric subdivision we have $\{x\} \subseteq \tau$ for all $\tau \in \sigma$. Recall that the vertices of $\overline{\sigma}$ are given by the vertices in σ. Thus, since $\sup p$ is a collection of vertices in $\overline{\sigma}$ we have that $\sup p \subseteq \sigma$. Then $\{x\} \subseteq \tau$ for all $\tau \in \sup p$, proving the lemma. $\square$

Now for the main result.

Theorem 3.1.4 (The Nerve Theorem)**.** Let $\mathcal{F} = \{F_i\}$ be a finite collection of closed, convex sets in $\mathbb{R}^N$. Then the geometric realization of the nerve of F and the union of all sets contained in $\mathcal{F}$ are homotopy equivalent: $\bigcup \mathcal{F} \simeq |\mathcal{N}(\mathcal{F})|$.

Proof. The idea of the proof is to construct two maps $f : \bigcup \mathcal{F} \to |\mathcal{N}(\mathcal{F})|$ and $g : |\mathcal{N}(\mathcal{F})| \to \bigcup \mathcal{F}$ so that their composition is homotopic to the identity.

More precisely, we will construct f so that it maps the areas in $\bigcup \mathcal{F}$ that are not an intersection, to the 0-simplices in $|\mathcal{N}(\mathcal{F})|$. Then, 1-fold intersections are mapped to 1-simplices, 2-fold intersections to 2-simplices and so on and so forth. Furthermore, g will be constructed by first mapping vertices σ of $\mathcal{B}(\mathcal{N}(K))$ into points in $\bigcup \mathcal{F}$ that lie in $\cap \sigma$. Then g is expanded to the whole geometric realization by linear interpolation.

[1] The double brackets around x signify that $\{\{x\}\}$ consists of a single vertex $\{x\} \in K$

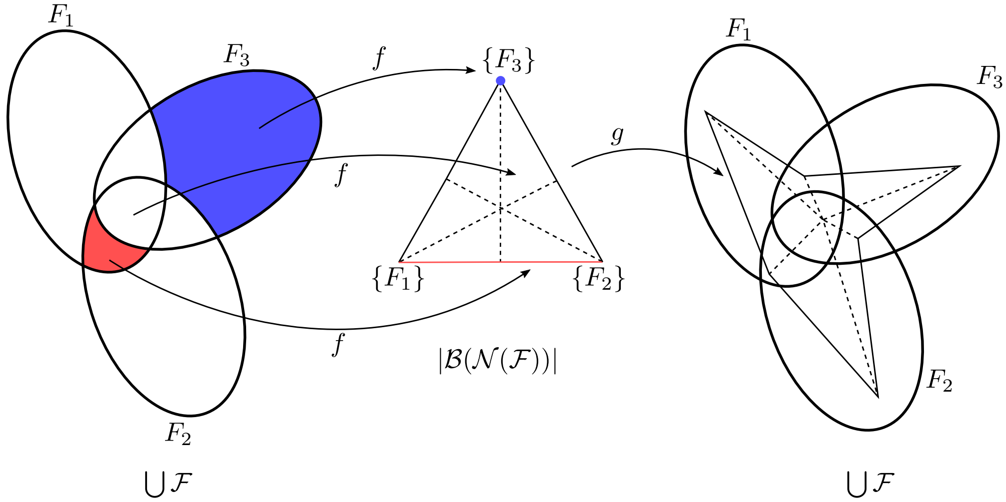

Figure 3.1: An example of how f and g can be constructed. The areas of no intersection in $\bigcup \mathcal{F}$ get mapped to the vertices of the triangle, the areas of 1-fold intersection get mapped to the edges and the 2-fold intersection maps to the interior of the triangle.

For each F_i we construct an open convex set $U_i \subseteq \mathbb{R}^N$ containing F_i by taking the union of open balls of radius $\varepsilon > 0$ around each point in F_i. That is:

$$U_i = \bigcup_{x \in F_i} \mathring{B}(x, \varepsilon)$$

Then clearly, $\mathcal{U} = \{U_i\}$ is a finite open cover of $\bigcup \mathcal{F}$. Since $\mathbb{R}^N$ is a normal space, $\bigcup \mathcal{F}$ admits a partition of unity subordinated to $\mathcal{U}$. That is, there exist continuous functions $f_i : \bigcup \mathcal{F} \to [0, 1]$ such that f_i is supported in U_i and for any point $x \in \bigcup \mathcal{F}$ we have $\sum_i f_i(x) = 1$. Then, we define a map $f : \bigcup \mathcal{F} \to |\mathcal{N}(\mathcal{F})|$ by:

$$f(x) = \sum_i f_i(x) F_i$$

where we take each F_i as an element in $|\mathcal{N}(\mathcal{F})|$ (recall the definition in 1.2.1). What we are doing here is that we are taking the function values of the partition of unity $\{f_i\}$ as barycentric coordinates of points in $|\mathcal{N}(\mathcal{F})|$.

Lets say f' is another such function, given by a different partition of unity $\{f_i'\}$. Let $x \in \bigcup \mathcal{F}$ and J the set of indices $j \in J$ such that $x \in U_j$. Then $f(x)$ and $f'(x)$ are both contained in $|\{ F_j \mid j \in J \}| \subseteq |\mathcal{N}(\mathcal{F})|$. Since the geometric realization of simplices are convex, the line joining $f(x)$ and $f'(x)$ is contained in $|\mathcal{N}(\mathcal{F})|$. Since for a given cover, the space of all its partitions of unity is convex, f and f' are homotopic by linear interpolation.

We now define a map $g : |\mathcal{N}(\mathcal{F})| \to \bigcup \mathcal{F}$ which will turn out to be a homotopy inverse of f. Taking $B := \mathcal{B}(\mathcal{N}(K))$ as the barycentric subdivision of the nerve, we will define g as a map from $|B|$ to $\bigcup \mathcal{F}$. By lemma 1.2.7, this defines g as a map on $|\mathcal{N}(\mathcal{F})|$. Recall that simplices $\{\sigma_0, ..., \sigma_n\} \in B$ are sets of simplices $\sigma_i \in \mathcal{N}(\mathcal{F})$.

We choose any mapping $\pi : \mathcal{N}(\mathcal{F}) \to \bigcup \mathcal{F}$ that takes a vertex $\{\sigma_n\} \in B$ to any point $x \in \bigcap \sigma_n$. We will call any such map a *projection*. Let $p \in |B|$ and recall that we can write $p = \sum_i \lambda_i \sigma_i$ with λ_i the barycentric coordinates of p. Heuristically, if we envision p as a point in $\bigcup \mathcal{F}$ we would like the support $\sup(p)$ of p to consist of all $\sigma_i \in \mathcal{N}(\mathcal{F})$ for which $p \in \bigcap \sigma_i$. This motivates the following definition of $g : |B| \to \bigcup \mathcal{F}$:

$$g(p) = \sum_{\sigma \in \sup(p)} \lambda_i \pi(\sigma)$$

Now let g' be another such function given by another projection $\pi' : B \to \bigcup \mathcal{F}$. We note that for any $\sigma \in \mathcal{N}(\mathcal{F})$ the points $\pi(\sigma)$ and $\pi'(\sigma)$ are both contained in $\bigcap \sigma_n$. Since $\bigcap \sigma_n$ is an intersection of convex sets, it is also convex so that the line connecting $\pi(b)$ and $\pi'(b)$ is contained in $\bigcap \sigma_n$. Thus, π and π' are homotopic by linear interpolation. Clearly then, g and g' are also homotopic.

We continue by proving that $f \circ g \simeq \mathrm{Id}$. Notice how we have not made any strict requirements for the partition of unity $\{f_i\}$ so that the map f it induces might as well send the whole of $\bigcup \mathcal{F}$ to a single vertex of $|\mathcal{N}(\mathcal{F})|$. What we did show however, is that any other partition of unity induces a function f' that is homotopic to f. For a given composition $f \circ g$, we will now construct a new function f' depending on g such that $f \circ g \simeq f' \circ g \simeq \mathrm{Id}$.

Writing $\mathrm{St}(F_i)$ for $\mathrm{St}(\{\{F_i\}\}) \subseteq B$, we will now show that the image of $|\overline{\mathrm{St}}(F_i))|$ under g is contained in U_i. Let $p \in |\overline{\mathrm{St}}(F_i))|$. By 3.1.3, each element in $\sup(p)$ contains F_i. Thus π send each element of $\sup p$ to a point in F_i and consequently $g(p)$ is now a linear combination of points in F_i. Since F_i is convex we have that $g(p) \in F_i \subseteq U_i$.

Furthermore, since there are only finitely many simplices in $\overline{\mathrm{St}}(F_i)$, the image of its geometric realization under g is compact, so it is closed and bounded. Therefore, there exists a convex open subset U_i' of U_i containing $g(|\overline{\mathrm{St}}(F_i))|$ such that $\overline{U_i'} \subseteq U_i$ and such that the U_i' still cover $\bigcup \mathcal{F}$. Note here that we can indeed construct the U_i' as convex by constructing it in the same way as U_i but for a radius $\varepsilon' < \varepsilon$. We can now choose a new partition of unity $\{f_i'\}$, still subordinated to $\mathcal{U}$, but with the extra condition that each f_i' is positive on $g(|\overline{\mathrm{St}}(F_i))|$ and $f_i' = 0$ outside of U_i'. This partition induces a map f' homotopic to f.

Now let p be a point in $|B|$. Since $|B| = \bigcup_i |\overline{\mathrm{St}}(F_i)|$, $p \in |\overline{\mathrm{St}}(F_k)|$ for some $F_k \in \mathcal{F}$. Since f_k' is positive on $g(|\overline{\mathrm{St}}(F_k)|)$, we have $f_k' \circ g(p) > 0$, giving $\{F_k\} \in \sup(f' \circ g(p)) \in B$. By the definition of the barycentric subdivision, $\{F_k\} \subseteq \tau$ for all $\tau \in \sup p$ so that by 3.1.3, $f' \circ g(p)$ is also contained in $|\overline{\mathrm{St}}(F_k)|$. Since the geometric realization of a star is convex, the line joining p and $f' \circ g(p)$ is contained in $\bigcup \mathcal{F}$ and thus $f' \circ g$ is homotopic to the identity.

To prove that $g \circ f' \simeq \mathrm{Id}$, we will proceed similarly. Let $x \in \bigcup \mathcal{F}$. Thus $x \in U_k$ for a certain k, giving $f_k'(x) > 0$ so that $\{F_k\} \in \sup(f'(x)) \in B$. Arguing similarly by the definition of the barycentric subdivision, each $\tau \in \sup(f'(x))$ contains $\{F_k\}$. Then by 3.1.3 $f'(x) \in |\overline{\mathrm{St}}(F_k)|$. But then $g \circ f'(x) \in g(\overline{\mathrm{St}}(F_k)) \subseteq U_k'$, so that x and $g \circ f'(x)$ are connected by a line contained in U_k'. So again, $g \circ f'$ is homotopic to the identity by linear interpolation. □

3.2 Čech Complex

Assume we have finite a point cloud $S \subseteq \mathbf{R}^n$, for example a set of data points, each with n variables. When we say we want to find the underlying *structure* of the point set S, we are assuming S is sampled from some specific subspace $E \subseteq \mathbb{R}^n$ and we would like to show topological properties of this space E. More precisely, we would like to recover the *homology* of E. Since S is a discrete set, it does not have any interesting topology on its own, but through *thickening* S by constructing balls around the points in S and looking at their union, we hope to obtain more structure. The collection of these balls induce a simplicial complex called the Čech complex.

Definition 3.2.1. For a finite set $S \subseteq \mathbb{R}^n$, its Čech complex $\mathcal{C}(S,\varepsilon)$ with radius $\varepsilon > 0$ is defined as the nerve of $S_\varepsilon := \{\ B_\varepsilon(x) \mid x \in S\ \}$. That is:

$$\mathcal{C}(S,\varepsilon) := \mathcal{N}(S_\varepsilon)$$

where S_ε is the collection of balls with radius ε, centered around points in S.

By the Nerve Theorem 3.1.4, we know that the geometric realization of $\mathcal{C}(S,\varepsilon)$ is homotopy equivalent to the union of all balls in S_ε. Since the homology of polyhedra is homotopy invariant, we see that we can capture the homology structure of S_ε by looking at the homology of the Čech complex.

Lemma 3.2.2. Given a point cloud $S \subseteq \mathbb{R}^N$, and $\varepsilon \leq \varepsilon'$, then $\mathcal{C}(S,\varepsilon) \subseteq \mathcal{C}(S,\varepsilon')$.

Proof. Each ball $B(x,\varepsilon) \subseteq B(x,\varepsilon')$. Thus, after identifying balls the same origin, any two balls in S_ε that intersect also intersect in $S_{\varepsilon'}$. ☐

Lemma 3.2.3. Given a point cloud $S \subseteq \mathbb{R}^N$, there is an ε_n such that $\mathcal{C}(S,\varepsilon) \subseteq \mathcal{C}(S,\varepsilon_n)$ for any $\varepsilon \leq \varepsilon_n$. We call $\mathcal{C}(S,\varepsilon_n)$ the **maximal Čech complex** of S.

Proof. Let $\varepsilon_n := \max\{\|x-y\| | x,y \in S\}$. Then for any $x \in S$, the ball $B(x,\varepsilon_n)$ contains S so that the intersection of any two balls in S_{ε_n} contains S as well. Thus $\mathcal{N}(S_{\varepsilon_n})$ equals the power set $\mathcal{P}(S_{\varepsilon_n})$. For any $\varepsilon \leq \varepsilon_n$, identifying balls with the same origin, we have that $\mathcal{N}(S_\varepsilon) \subseteq \mathcal{P}(S_{\varepsilon_n}) = \mathcal{N}(S_{\varepsilon_n})$ as had to be proven. ☐

Since S is finite, the different ε for which a new simplex appears in $\mathcal{C}(S,\varepsilon)$ forms a discrete set $\varepsilon_1, ..., \varepsilon_n$. Thus any two Čech complexes associated with radii contained in a single interval $[\varepsilon_i, \varepsilon_{i+1})$ are isomorphic. Thus it makes sense to write all the Čech complexes of a point cloud S as finite sequence

$$0 = K_0 \subseteq K_1 \subseteq ... \subseteq K_n = K$$

where $K_i := \mathcal{C}(S,\varepsilon)$ and where each K_i is a subcomplex of all the larger complexes in the chain.

We can define a function $f : K \to \mathbb{R}$ that maps each simplex in K to the ε_i for which it first appears in $\mathcal{C}(S,\varepsilon)$ so that the subcomplexes K_i shown above can now be seen as sublevel sets of $K_i := f^{-1}(-\infty, \varepsilon_i]$ of f. We can generalize this construction to any map from a simplicial complex K to $\mathbb{R}$ to get something we call a filtration.

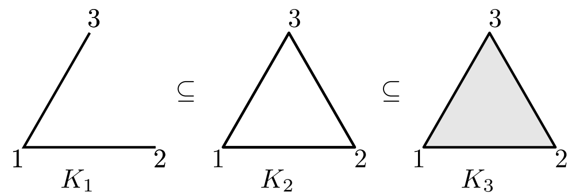

Figure 3.2: A chain of simplicial complexes K_1, K_2, K_3.

3.3 Filtrations

Definition 3.3.1. Let K be a simplicial complex and $f : K \to \mathbb{R}$ a *monotonic* map, assigning to each *simplex* of K a value in $\mathbb{R}$, such that if τ is a face of σ, then $f(\tau) \leq f(\sigma)$. The **filtration** of f is the collection:

$$\{K(\varepsilon) := f^{-1}(-\infty, \varepsilon] \,|\, \varepsilon \in \mathbb{R}\}$$

In this context, the function f is called the *filter* and $\varepsilon \in \mathbb{R}$ is called the *filtration parameter*.

Since K is finite, $f(K) \subseteq \mathbb{R}$ is also finite. Thus we see that the filtration

$$\{\, K(\varepsilon) = f^{-1}(-\infty, \varepsilon] \mid \varepsilon \in \mathbb{R} \,\}$$

only contains finitely many distinct sets. We can index the elements in $f(K)$ as $\varepsilon_0 < \varepsilon_1 < ... < \varepsilon_n$ and by extension we can index each element in the filtration by $K_i := K(\varepsilon_i)$. Thus a filtration can be seen as a chain:

$$K_0 \subseteq K_1 \subseteq ... \subseteq K_n$$

where n is not necessarily equal to the cardinality of K, as simplices in K might get mapped to the same value in $\mathbb{R}$. We see that this construction is not unique to the Čech complex. All we need is a filter $f : K \to \mathbb{R}$ on a simplicial complex K to obtain a filtration. Note however that n is not necessarily equal to the cardinality of K, as simplices in K might get mapped to the same value in $\mathbb{R}$.

Of course we can also consider the homology groups of each simplicial complex in the filtration. The inclusions in a chain of simplicial complexes $K_i \hookrightarrow K_j$ induce homomorphisms $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ so that we get, for each p, a chain of homology groups connected by homomorphisms:

$$0 \xrightarrow{f_p^{0,1}} H_p(K_1) \xrightarrow{f_p^{1,2}} ... \xrightarrow{f_p^{n-1,n}} H_p(K_N)$$

In each step along this chain, homology classes might appear or disappear.

Example 3.3.2. Take for example the three simplicial complexes shown in figure 3.2 that form a filtration of some $f : K := K_3 \to \mathbb{R}$. We see that the class $\{1,2\} + \{2,3\} + \{3,1\}$ first appears in $H_1(K_2)$ by the addition of the 1-simplex $\{2,3\}$ in K_2. In the subsequent complex K_3 it disappears in the its first homology class, by the addition of the 2-simplex $\{1,2,3\}$. Indeed: $\{1,2\} + \{2,3\} + \{3,1\} = \partial_2(\{1,2,3\}) \subseteq \mathrm{Im}(\partial_2)$. $\triangle$

In applications, we would like to study the structure of our simplicial complex using homology. Since the homology changes as we change the filtration parameter ε a natural question to ask is: *which ε should we pick to best detect the feature that we are interested in?* The problem is that in a lot of cases there might not be an a priori right choice for ε. Instead of trying to find some procedure or rationale for picking ε, we will discuss a different idea. We will study a representation of the homology structure underlying S, without introducing any restraints on the filtration parameter.

3.4 Persistence Modules

Instead of asking ourselves which filtration parameters ε to use, we will look at how the homology of a filtration changes as we change ε. In particular, we study those classes that *persist* over a longer range of ε's, and collect these in groups called *persistence homology groups*.

First we will more precisely define what we mean by the homology structure of a point set S through the notions of the *persistence module* and its associated *persistent homology*.

Then we will study a useful representation of the persistence module called the *persistence diagram*. We will define the *bottleneck distance* which is a metric on the space of all persistence diagrams. Lastly we will see that under this metric the persistence diagram is stable with respect to perturbations in the filtration. That is, small changes in our data only result in small changes in the persistence diagram. We will follow [8, Chapter 1] and [2, Section 7.2].

In the following section we will fix a $p \in \mathbb{N}$ and write $H(K)$ for the p-th homology vector space $H_p(K)$.

Definition 3.4.1. Given a filtration

$$\left\{ f^{-1}(-\infty, x] \;\middle|\; x \in \mathbb{R} \right\}$$

of a monotonic function $f : K \to \mathbb{R}$, a **persistence module** of f is a collection:

$$V := \{H_x := H(f^{-1}(-\infty, x] \,|\, x \in \mathbb{R}\}$$

of homology vector spaces together with the collection of linear maps $f_x^y : H_x \to H_y$, where f_x^y for given $x < y$ is the function induced by the inclusion $f^{-1}(-\infty, x] \subseteq f^{-1}(-\infty, y]$.

So, the persistence module not only contains the information from the homology groups H_x, but also describes how homology classes change under the maps f_x^y. There is a useful way of aggregating this information by defining the *persistent homology spaces*.

Definition 3.4.2. Given a persistence module V, we define its **persistent homology spaces** H_x^y as the images of the maps f_x^y. That is:

$$H_x^y := \operatorname{Im} f_x^y \subseteq H_y$$

For each such vector space, we define its **persistent Betti number** as $\beta_x^y := \dim H_x^y$. By convention, we say that $H_x^y = 0$ when y is infinite.

Whenever we need to emphasize the dependence of the vector spaces H_x and H_x^y on f we will write $H_x(f)$ and $H_x^y(f)$ respectively.

We introduce some vocabulary for the homology groups of the filtration.

Definition 3.4.3. Given a filtration of $f : K \to \mathbb{R}$, we say that a class $\gamma \in H_x$ is *born* at x when $\gamma \notin H_{x-1}^{x}$. If γ was born at x, then it *dies* at y for some $y \geq x$ when it has merged with an older class, that is: $f_x^y(\gamma) \in H_{x-1}^{y}$, *and* y is the parameter where this is the case: $f_x^y(\gamma) \notin H_{x-1}^{y-1}$.

As stated before, we can index the distinct subcomplexes in a filtration by $K_i := f^{-1}(-\infty, \varepsilon_i]$ so that we might as well say that a class is born or dies in K_i, by which we mean it was born or died at ε_i.

Example 3.4.4. If we go back to example 3.3.2, we see that the class $\gamma := \{1,2\}+\{2,3\}+\{3,1\}$ is *born* in K_2. Then since $\{1,2\} + \{2,3\} + \{3,1\} \equiv 0$ as homology class in $H_p(K_3)$, $\gamma \in H_p^{1,3}$. In other words, γ merges with the trivial class and thus dies in K_3. $\triangle$

3.5 Persistence Diagrams

To visualize this information, we will now define a representation of the persistence homology called the *persistence diagram.* The persistence diagram will consist of the birth-times and death-times of all the homology classes in the filtration. Since at each step in the filtration, multiple classes might be born and die at the same time, the persistence diagram will end up being a multiset in $\overline{\mathbb{R}}^2 := \mathbb{R} \times (\mathbb{R} \cup \{-\infty, \infty\})$.

Definition 3.5.1. A **multiset** is a tuple (D, m) with D a set called the *underlying set* and $m : D \to \mathbb{N} \cup \{\infty\}$ a function called the *multiplicity.*

In other words, a multiset D is a set where for each $a \in D$ where $m(a)$ is the number of times that a appears in D.

We are now ready to define the persistence diagram.

Definition 3.5.2. Let $f : K \to \mathbb{R}$ be a monotonic function on a simplicial complex K. Let $\mathrm{Im}(f) = \{a_1, a_2, ..., a_n\}$ and let $b_0, ..., b_n$ be an *interleaved* sequence. That is:

$$b_{i-1} < a_i < b_{i-1}$$

Additionally, we define $b_{-1} = a_0 = -\infty$ and $b_{n+1} = a_{n+1} = \infty$. For the diagonal in $\mathbb{R}^2$ we use the notation $\Delta := \{ (x, x) \mid x \in \mathbb{R} \}$.

We consider the persistence module V of f consisting of the p-th homology spaces H_x. Then the p-th **persistence diagram** of f is the multiset $D(f) = \{ (a_i, a_j) \mid 0 \leq i < j \leq n+1 \} \cup \Delta \subseteq \overline{\mathbb{R}}^2$ with multiplicity defined as:

$$\mu(a_i, a_j) := \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}}$$

on the points (a_i, a_j) and as ∞ on the diagonal Δ. Where β_x^y are the persistent Betti-numbers of the persistence module V.

We include the diagonal Δ for technical reasons that will become clear in section 3.6. Furthermore, for any point $(a, b) \in D(f)$, we have $a \leq b$ so that $D(f) \subset \overline{\mathbb{R}}^2$ will lie in the region $\{ (x, y) \in \mathbb{R}^2 \mid y \geq x \}$ above the diagonal.

Since H_x^y only changes at values a_i, we define the multiplicity at (a_i, a_j) by looking at the persistent Betti-numbers at the interleaved values $b_i, b_{i-1}, b_j, b_{j-1}$ around it. For example, we interpret $\beta_{b_i}^{b_{j-1}}$ as the number of classes that are born *before* b_i and die *after* b_{j-1}. Then, $\mu(a_i, a_j)$ represents the number of classes that are born *between* b_i and b_{i-1} and die *between* b_j and b_{j-1}.

Remark 3.5.3. By the above interpretation, $(a_i, a_{n+1}) = (a_i, \infty)$ corresponds to classes that are born at a_i, but die at ∞. To make this precise, we could have extended $f : K \to \mathbb{R}$ as follows.

We construct $\overline{K}$ from K by adding a single vertex C and for each simplex $\sigma \in K$, adding $C \cup \mathcal{P}(\sigma)$. This turns out te be a simplicial version of taking the *cone* of a topological space and, similarly to the general case, $\overline{K}$ will have trivial homology. We take a new function $\overline{f} : \overline{K} \to \mathbb{R} \cup \{\infty\}$ which is defined as f on K and as $\sigma \mapsto \infty$ on $\overline{K} \setminus K$.

Instead of defining the persistence diagram for f, we could have defined it similarly for $\overline{f}$. Now, classes that survive the whole filtration of $f : K \to \mathbb{R}$, will become trivial in $\overline{K}$. Since $\overline{f}(\overline{K}) = \infty$, we say that this class dies at infinity.

Recall that in definition 3.4.2, we defined $H_x^y = 0$ whenever y is equal to infinity. The above discussion justifies this convention, since now $H_\infty = H(\overline{f}^{\,-1}(-\infty, \infty]) = H(\overline{K}) = \{0\}$.

We can plot all points in $D(f)$ on a diagram to obtain a graphic representation.

Example 3.5.4. Consider the filtration depicted in figure 3.3. We can consider the zeroth, first and second homology classes of the filtration as we move along the subcomplexes. For example, K_1 contains two connected components, so that the dimension of $H_0(K_1)$ equals two. One of these classes dies in the next subcomplex K_1 when the two components merge with each other. This means that the zeroth persistence diagram is the multiset $\{(1, 2), (1, \infty)\}$. We can plot this multiset in a 2-dimensional graph by a little trick. We take the largest $\varepsilon \in \operatorname{Im} f$ and draw a horizontal line at height ε that now represents $\{(x, \infty \mid x \in \mathbb{R}\}$. If we similarly do this for the first and second persistence diagram, we obtain the picture in figure 3.4. △

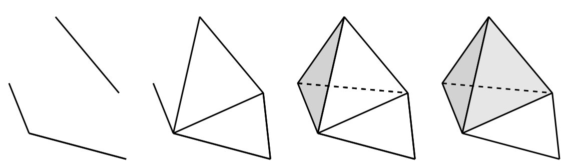

Figure 3.3: A filtration K_1, K_2, K_3, K_4.

We can now measure how *persistent* a homology class γ is by looking at how long it 'survives'.

Definition 3.5.5. Let $f : K \to \mathbb{R}$ define a filtration of a simplicial complex K. Let $\gamma \in H_x$ that is born at x and dies at y. The **persistence** of γ is the difference:

$$x - y$$

If γ does not die at any time y, then the persistence of γ is said to be ∞.

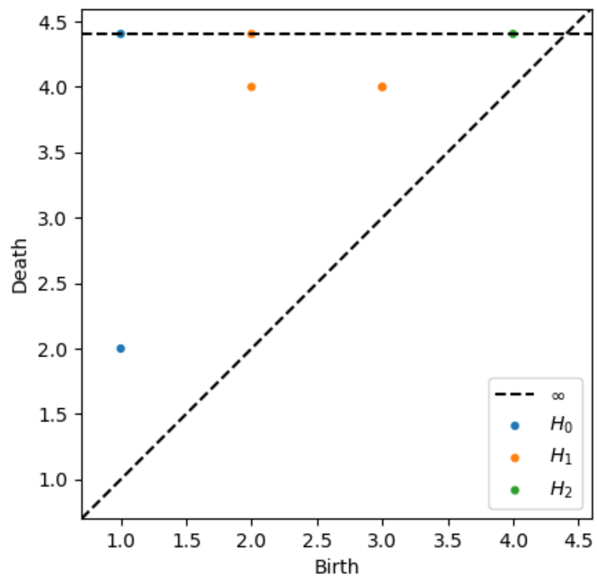

Figure 3.4: The persistence diagram obtained from the filtration depicted in figure 3.3.

So in our persistence diagram depicted in figure 3.4, the persistence of a class is depicted by its horizontal distance from the diagonal.

As it stands, the persistent homology spaces H_x^y of a monotonic function f give a corresponding persistence diagram $D(f)$. What we will now see is that the converse is also true. Each persistence diagram gives corresponding persistent homology spaces. We will do this by not just considering the multiplicity of a single point, of a whole region of the diagram.

Definition 3.5.6. The *multiplicity* of a subset R in a multiset (D, m) is defined as the sum of the multiplicities of elements in R. That is:

$$\#(R) := \sum_{\substack{(x,y)\in R \\ x\leq y}} m(x,y)$$

Using this, we define the *total multiplicity* of a persitence diagram.

Definition 3.5.7. The **total multiplicity** of a persistence diagram $D(f)$ is defined as

$$\#(D(f) - \Delta)$$

That is, we consider the multiplicity of the diagram, excluding the points on the diagonal Δ.

By the way we constructed the persistence diagram, each upper left quadrant $Q(x, y) := \{\, (p, q) \in D(f) \mid p \leq x, q \geq y \,\}$ corresponds to a persistent homology vector space.

Lemma 3.5.8 (k-Triangle Lemma, [3])**.** Let $f : K \to \mathbb{R}$ be monotonic and $x < y$ be real numbers such that (x, y) is not in the image of f. Then, the total multiplicity of the upper left quadrant $Q(x, y) \subseteq \overline{\mathbb{R}}^2$ of $D(f)$ is:

$$\#(D(f) \cap Q(x, y)) = \beta_x^y$$

Proof. We use the same definition as in 3.5.2 for the interleaved sequences $\{a_i\}_{i=0}^{n+1}$ and $\{b_i\}_{i=-1}^{n+1}$. Since the sublevel sets of f stay the same on the open intervals a_i, a_{i+1}, we can assume that $x = b_k$ and $y = b_{l-1}$ for some integers k, l. Since $x < y$, we also know that $k < l$. Then the multiplicity of $D(f) \cap Q(b_k, b_{l-1})$ will be a sum over all $(a_i, a_j) \in D(f)$ where $i \leq k$ and $l \leq j$. Since $i < j$ by definition of the persistence diagram, we can summarize these inequalities as $i \leq k < l \leq j$. We obtain the following sum:

$$\begin{aligned}
\#(D(f) \cap Q(b_k, b_{l-1})) &= \sum_{i \leq k < l \leq j} \mu(a_i, a_j) \\
&= \sum_{i \leq k < l \leq j} \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}} \\
&= \sum_{i \leq k} \beta_{b_{i-1}}^{b_{n+1}} - \beta_{b_i}^{b_{n+1}} + \beta_{b_i}^{b_{l-1}} - \beta_{b_{i-1}}^{b_{l-1}} \\
&= \beta_{b_{-1}}^{b_{n+1}} - \beta_{b_k}^{b_{n+1}} + \beta_{b_k}^{b_{l-1}} - \beta_{b_{-1}}^{b_{l-1}}
\end{aligned}$$

In the last two equalities, we use the fact that $\sum \mu(a_i, a_j)$ is a telescoping sum when we fix one of the two indices. Recall that we defined $b_{-1} = -\infty$ and $b_{n+1} = \infty$. By a slight abuse of notation: $H_{-\infty} = H(f^{-1}(-\infty, \infty]) = \{0\}$, so $\beta_{b_{-1}}^{b_{l-1}} = 0$. Additionally we said that $H_x^y = 0$ whenever y is infinite. Thus, all terms in the last expression above are equal to 0, except for $\beta_{b_k}^{b_{l-1}} = \beta_x^y$. This proves the lemma. □

Since the persistent homology spaces H_x^y are finite dimensional, they are completely determined by their dimension β_x^y. Thus, whenever we have a persistence diagram, we find its corresponding persistent homology spaces by considering the total multiplicities in the upper left quadrants of $D(f)$.

We see that upper-left quadrants $Q(x, y)$ in the persistence diagram $D(f)$ correspond to all classes that were born before x and died after y and their total multiplicity is equal to the dimension of H_x^y, see figure 3.5. We can do even better, by relating not just quadrants but squares to specific vector spaces. To do this, we need some more definitions.

Definition 3.5.9. Let $f : K \to \mathbb{R}$ be a monotonic map and $w < x < y < z$ not in the image of f. We define the following two subspaces of H_x^y.

$H_x^{y,z}$ Since the inclusions $f^{-1}(-\infty, x] \subseteq f^{-1}(-\infty, y] \subseteq f^{-1}(-\infty, z]$ commute, the diagram in figure 3.6 also commutes. Thus, the restriction of f_y^z to $H_x^y = \mathrm{Im}(f_x^y)$ defines a surjection $f_x^{y,z} : H_x^y \to H_x^z$. We then define $H_x^{y,z} := \ker(f_x^{y,z}) \subseteq H_x^y$.

$H_{w,x}^{y,z}$ Since $H_w^y = \mathrm{Im}(f_w^y) = \mathrm{Im}(f_x^y \circ f_w^x) \subseteq \mathrm{Im}(f_x^y) = H_x^y$, we can write $f_w^{y,z}$ as the restriction of $f_x^{y,z}$ to H_w^y. Thus $\ker(f_w^{y,z}) \subseteq \ker(f_x^{y,z})$ so that the quotient $H_{w,x}^{y,z} := H_x^{y,z} / H_w^{y,z}$ is well-defined.

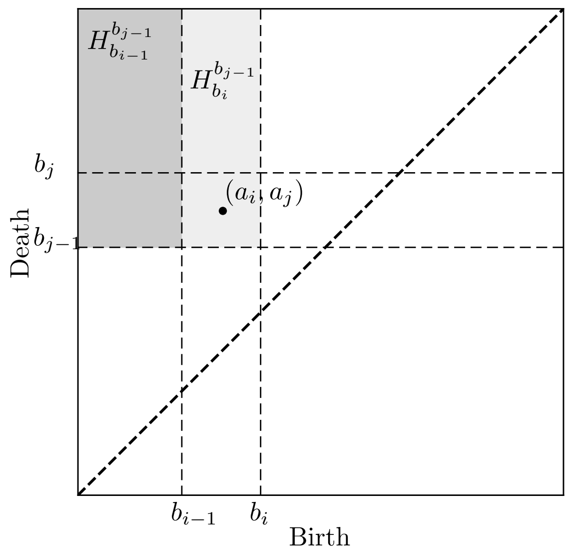

Figure 3.5: The greyed-out region contains all points representing classes that are born before b_i and die after b_{j-1}. Similarly, the dark-grey region corresponds to classes that are born before b_{i-1} and die after b_{j-1}. The light-grey area then corresponds to classes that die after b_{j-1}, but we born *between* b_{i-1} and b_i.

$$\begin{tikzcd} H_x \arrow[r, "f_x^y"] \arrow[dr, "f_x^z"'] & H_y \arrow[d, "f_y^z"] \\ & F_z \end{tikzcd}$$

Figure 3.6: Commutative diagram of the maps between H_x, H_y, H_z induced by the inclusions between level sets of f.

Similar to how we do this with persistent homology spaces, we will write $H_x^{y,z}(f)$ and $H_{w,x}^{y,z}(f)$ whenever we want to emphasize the dependence of the spaces on f.

Now, we apply the k-Triangle Lemma (3.5.8) to relate regions in the persistence diagram $D(f)$ to the vector spaces defined above.

Lemma 3.5.10. Let $f : K \to \mathbb{R}$ be monotonic on a simplicial complex K. Let $w < x < y < z$ be elements in $\mathbb{R} \setminus \mathrm{Im}(f)^2$. Then the following statements are true:

1. $\dim H_x^y = \#(D(f) \cap [-\infty, x] \times [y, \infty])$.
2. $\dim H_x^{y,z} = \#(D(f) \cap [-\infty, x] \times [y, z])$.
3. $\dim H_{w,x}^{y,z} = \#(D(f) \cap [w, x] \times [y, z])$.

Proof. We write $Q(x, y) := [-\infty, x] \times [y, \infty] \subseteq \overline{\mathbb{R}}^2$.

[2]This restriction on w, x, y, z ensures that we can apply 3.5.8

1. This is just the k-Triangle Lemma (3.5.8), repeated here for completeness.
2. Recall that $\operatorname{Im} f_x^{y,z} = H_x^z$ and $\ker f_x^{y,z} = F_x^{y,z}$. Then, from linear algebra and using the above item, we know that $\dim H_x^{y,z} = \dim H_x^y - \dim H_x^z = \beta_x^y - \beta_x^z = \#(D(f) \cap (Q(x,y) \setminus Q(x,z))) = \#(D(f) \cap ([-\infty, x] \times [y,z]))$.
3. This follows similarly by using the dimension formula for quotients and using the above item.

□

The power of this lemma lies in the fact that we can now describe specific 'boxes' in the persistence diagram $D(f)$ by specific vector spaces. In the next section, we will use this tool extensively to relate persistence diagrams induced by distinct maps $f, g : K \to \mathbb{R}$ to each other through vector spaces that are determined by f and g.

3.6 Bottleneck Stability

In the context of data analysis, we are given a dataset $X \subseteq \mathbb{R}^n$ in which we try to differentiate *features* from *noise*. Heuristically, noise consists of small variations in the data that do not exhibit any particular pattern that we are trying to detect.

Our hope is, that using persistence diagrams, we will able to recognize these small variations as homology classes that have a very short persistence. That is, if we have two datasets X and Y measuring the same phenomenon, but differing slightly because of noise, we hope that their associated persistence diagrams will also only differ slightly. To see this, we first define the *bottleneck distance*.

We can see a persistence diagram $D(f)$ as a countably infinite set by interpreting elements $p \in D(f)$ with multiplicity $\mu(p) > 1$ as $\mu(p)$ distinct elements. In the case of multiplicity ∞, we consider a countably infinite set of added distinct elements, so that the whole set remains countable. Having made these adjustments, we can now relate two persistence diagrams by looking at all bijections between them.

Definition 3.6.1. Let X and Y be two persistence diagrams. We define the **bottleneck distance** between X and Y as:

$$W_\infty(X, Y) = \inf_{\phi} \sup_{x \in X} \|x - \gamma(x)\|_\infty$$

where $\phi : X \to Y$ goes over all bijections between X and Y and $\|(x, y)\|_\infty = \max(|x|, |y|)$ is the L_∞-norm on $\mathbb{R}^2$.

Recall that we defined persistence diagrams to include all points on the diagonal with infinite multiplicity. Thus, persistence diagrams are always countably infinite so that we can always find bijections between them.

The subject of this section will be the following theorem. The original proof in [1], will be explained here in detail.

Theorem 3.6.2 (Bottleneck Stability [1])**.** Let $f, g : K \to \mathbb{R}$ be monotonic functions. Then:

$$W_\infty(D(f), D(g)) \leq \|f - g\|_\infty$$

This means that if two filtrations are similar, then their persistence diagrams will also be very similar. In other words, the persistence diagram is a *stable* representation of the filtration. We can therefore use the persistence diagram in data analysis, as small uncertainty in the point cloud and consequently a small uncertainty in the filtration, will only result in a small change in the persistence diagram.

First we will prove a particular case of the theorem called the *Easy Bijection Lemma*, where we assume a certain restriction on the distance $\|f - g\|_\infty$. We do this by constructing a particular bijection between the diagrams $D(f)$ and $D(g)$. The idea is to relate the multiplicity of a rectangle in $D(g)$ to the multiplicity of a slightly smaller rectangle in $D(f)$. If we take the rectangle $R \subseteq D(g)$ small enough, using the *Box Lemma* its multiplicity will correspond to the multiplicity of a single point in $D(f)$ so that we can coherently map R into this point.

To prove the general theorem then, we interpolate between f and g by maps that are close enough so that the Easy Bijection Lemma applies.

Before we prove the Box Lemma, we need the following result.

Lemma 3.6.3. Given two monotonic functions $f, g : K \to \mathbb{R}$, their persistence diagrams V, W, and writing $\varepsilon := \|f - g\|_\infty$, we have that the V and W are ε-**interleaved**. That is, for all $b < c$ there exists a map $\psi_c : H_c(g) \to H_{c+\varepsilon}(f)$ such that:

$$H_{b-\varepsilon}^{c+\varepsilon}(f) \subseteq \psi_c(H_b^c(g)) \subseteq H_{b+\varepsilon}^{c+\varepsilon}(f) \tag{3.6.1}$$

Proof. Since $|f(\sigma) - g(\sigma)| \leq \varepsilon$ we have that $g(\varepsilon) \leq f(\sigma) + \varepsilon \leq x + \varepsilon$. So $f^{-1}(-\infty, x] \subseteq g^{-1}(-\infty, x + \varepsilon]$. Let $\phi_x : H_x(f) \to H_{x+\varepsilon}(g)$ be the map induced by this inclusion. The symmetric inclusion also holds, and the map it induces we call $\psi_x : H_x(g) \to H_{x+\varepsilon}(f)$.

Combining these maps with the one we already defined in 3.5.9 and writing $F_x := H_x(f)$ and $G_x := H_x(g)$ we get the following diagram.

$$\begin{tikzcd}
F_{b-\varepsilon} \arrow[r, "f_{b-\varepsilon}^{c+\varepsilon}"] \arrow[d, "\phi_{b-\varepsilon}"] & F_{c+\varepsilon} & F_{b+\varepsilon} \arrow[r, "f_{b-\varepsilon}^{c+\varepsilon}"] & F_{c+\varepsilon} \\
G_b \arrow[r, "g_b^c"] & G_c \arrow[u, "\psi_c"'] & G_b \arrow[r, "g_b^c"] \arrow[u, "\phi_b"'] & G_c \arrow[u, "\psi_c"']
\end{tikzcd}$$

Since all the maps here are induced by inclusions, and since inclusions commute, this diagram also commutes. From the diagram we now see that $f_{b-\varepsilon}^{c+\varepsilon} = \psi_c \circ g_b^c \circ \phi_{b-\varepsilon}$ so clearly:

$$F_{b-\varepsilon}^{c+\varepsilon} = \operatorname{Im} f_{b-\varepsilon}^{c+\varepsilon} = \operatorname{Im}(\psi_c \circ g_b^c \circ \phi_{b-\varepsilon}) \subseteq \psi_c(\operatorname{Im} g_b^c) = \psi_c(G_b^c)$$

Now for the other inclusion, we turn to the right diagram which by commutativity gives $\psi_c \circ g_b^c = f_{b+\varepsilon}^{c+\varepsilon} \circ \psi_b$. Thus:

$$\psi(G_b^c) = \operatorname{Im} \psi_c \circ g_b^c = \operatorname{Im} f_{b+\varepsilon}^{c+\varepsilon} \circ \psi_b \subseteq \operatorname{Im} f_{b+\varepsilon}^{c+\varepsilon} = F_{b+\varepsilon}^{c+\varepsilon}$$

which proves the lemma. $\square$

The following lemma will a crucial step in proving stability.

Lemma 3.6.4 (Box Lemma, [1])**.** Let $f, g : K \to \mathbb{R}$ be two monotonic functions on a simplicial complex K. For real numbers $a < b < c < d$ and $\varepsilon := \|f - g\|_\infty$, we consider the rectangle $R = [a, b] \times [c, d]$, and the rectangle $R_\varepsilon = [a + \varepsilon, b - \varepsilon] \times [c + \varepsilon] \times [d - \varepsilon]$ obtained by shrinking R. Then:

$$\#(D(f) \cap R_\varepsilon) \leq \#(D(g) \cap R)$$

Proof. Before we begin the proof, we introduce some short-hand notation. We will write $G_x := H_x(g)$ and $F_x := H_x(f)$ as in lemma 3.6.3. Furthermore we write $F_x := H_x(f)$,$F_x^{y,z} := H_x^{y,z}(f)$ and $F_{w,x}^{y,z} := H_{w,x}^{y,z}(f)$ and similarly for the spaces associated with g.

Applying the k-Triangle Lemma. We reduce the problem to the case that $a, b, c, d \notin \operatorname{Im} g$ and $a + \varepsilon, b - \varepsilon, c + \varepsilon, d - \varepsilon \notin \operatorname{Im} f$. Assuming otherwise, since $D(g) \subseteq \operatorname{Im} f$ it is a discrete finite set. Thus we can make R slightly larger so that $D(g) \cap R$ does not change. Similarly, this is true for R_ε and $D(f)$. This means that we can apply the k-Triangle Lemma 3.5.8 to see that:

$$\dim F_{a+\varepsilon,b-\varepsilon}^{c+\varepsilon,d-\varepsilon} = \#(D(f) \cap R_\varepsilon) \tag{3.6.2}$$

$$\dim G_{a,b}^{c,d} = \#(D(g) \cap R) \tag{3.6.3}$$

Thus to prove the lemma, we need to show that $F_{a+\varepsilon,b-\varepsilon}^{c+\varepsilon,d-\varepsilon}$ is a subspace of $G_{a,b}^{c,d}$. We would hope to be able to prove this by showing that its constituent homology spaces of the form $F_{x\pm\varepsilon}^{y\pm\varepsilon}$ are subspaces of G_x^y and vice versa. The interleaving lemma 3.6.3 helps us with this, but there is one problem: it does not tell us that G_b^c is a subspace of $F_{b-\varepsilon}^{c+\varepsilon}$ and in fact this is not generally the case.

Constructing the diagram. Instead our strategy will be to define some *subspace* $E_{a,b}^{c,d} \subseteq G_{a,b}^{c,d}$. Then, we hope to be able to show that$F_{a+\varepsilon,b-\varepsilon}^{c+\varepsilon,d-\varepsilon}$ is a subspace of this $E_{a,b}^{c,d}$ by which we would have probed the lemma. To this end we will construct the diagram shown in 3.7. We let $u_2 = f_{a+\varepsilon}^{c+\varepsilon,d-\varepsilon}$ and $u_3 = f_{b-\varepsilon}^{c+\varepsilon,d-\varepsilon}$. Now assuming we have defined E_a^c as some subspace of G_a^c, the maps s_1 and s_2 are given by the interleaving due to lemma 3.6.3. That is: s_1 is the restriction of $\phi_{b-\varepsilon}$ to $F_{b-\varepsilon}^{d-\varepsilon}$ and s_2 is the restriction of ψ_a to E_a^c and its image is guaranteed to be in the codomain by lemma 3.6.3.

Similarly, we would like s_3 to be some restriction of the map ψ_c, but a priori, its image is not guaranteed to be in $F_{b-\varepsilon}^{c+\varepsilon}$. By 3.6.3 we at least know that:

$$F_{b-\varepsilon}^{c+\varepsilon} \subseteq \psi_c(G_b^c) \tag{3.6.4}$$

So firstly, we need to define E_b^c as some subspace of G_b^c so that its image under ψ_c is contained in $F_{b-\varepsilon}^{c+\varepsilon}$ and secondly, for reasons that will become clear in a moment, we want the image of s_3 to be equal to $\ker u_3 = F_{b-\varepsilon}^{c+\varepsilon,d-\varepsilon}$. Therefore we define E_b^c as the preimage of $\ker u_3$ under ψ_c, and by 3.6.4 the *whole* preimage is indeed contained in G_b^c. We can now define $E_a^c = E_b^c \cap G_a^c$.

Taking the images of the maps in diagram 3.6 we see that the domains of the functions r_i are contained in their codomain. Thus we simply define r_i as the inclusion maps. Lastly, we

$$
\begin{tikzcd}
G_a^d \arrow[rrr, "r_1"] & & & G_b^d \\
& F_{a+\varepsilon}^{d-\varepsilon} \arrow[r, "r_2"] & F_{b-\varepsilon}^{d-\varepsilon} \arrow[ur, "s_1"] & \\
& F_{a+\varepsilon}^{c+\varepsilon} \arrow[u, "u_2"] \arrow[r, "r_3"] & F_{b-\varepsilon}^{c+\varepsilon} \arrow[u, "u_3"'] & \\
G_a^c \subseteq E_a^c \arrow[uuu, "u1"] \arrow[ur, "s_2"] \arrow[rrr, "r_4"] & & & E_b^c \subseteq G_b^c \arrow[uuu, "u4"'] \arrow[ul, "s_3"]
\end{tikzcd}
$$

Figure 3.7: Commutative diagram of the homology vector spaces associated with f and g. Recall that each vector space on the corner of the two boxes corresponds to a quadrant of the corresponding persistence diagram. For example: $F_{a+\varepsilon}^{d-\varepsilon}$ corresponds to $D(f) \cap [-\infty, a+\varepsilon] \times [d-\varepsilon, \infty]$. In this way, the corners correspond to corners of the boxes R_ε and R.

define u_1 as the restriction of $g_a^{c,d}$ to E_a^c and similarly we define u_4 as the map $g_b^{c,d}$ restricted to E_b^c. Since we defined each map in the diagram as either an inclusion, or as the restriction of a function induced by an inclusion, the diagram commutes.

Defining $E_{a,b}^{c,d} \subseteq G_{a,b}^{c,d}$. Recall that we defined $G_{a,b}^{c,d}$ by taking the quotient of kernels of maps between G_b^c and G_a^c. Now we will do the same for the subspaces $E_b^c \subseteq G_b^c$ and $E_a^c \subseteq G_a^c$ to get our vector space $E_{a,b}^{c,d}$.

First, we see that $u_4 = s_1 \circ u_3 \circ s_3$, but since $\operatorname{Im} s_3 = \ker u_3$ and s_1 is an inclusion: $u_4(E_b^c) = 0$. Analogous to how we defined the vector spaces $H_x^{y,z}$ we take $E_b^{c,d} := E_b^c = \ker u_4$. We can do a similar thing for E_a^c by the equality $r_1 \circ u_1 = u_4 \circ r_4$. Since r_1 and r_4 are inclusions and by the definition of E_a^c we have $u_1(E_a^c) = u_4(E_a^c) \subseteq u_4(E_b^c) = 0$. Having done this we define $E_a^{c,d} := E_a^c = \ker u_1 =$. We are now ready to define:

$$
E_{a,b}^{c,d} := E_b^{c,d} / E_a^{c,d} = \ker u_4 / \ker u_1
$$

To show that $E_{a,b}^{c,d}$ is indeed a subspace of $G_{a,b}^{c,d}$ we prove:

$$
\dim F_{a+\varepsilon, b-\varepsilon}^{c+\varepsilon, d-\varepsilon} \leq \dim E_{a,b}^{c,d} \leq \dim G_{a,b}^{c,d} \tag{3.6.5}
$$

by constructing an injection from $E_{a,b}^{c,d}$ to $G_{a,b}^{c,d}$.

We define the map $E_{a,b}^{c,d} \to G_{a,b}^{c,d}$ by sending each equivalence class $x + E_a^{c,d}$ to the equivalence class $x + G_a^{c,d}$. To show this is well-defined, let x, y be equivalent classes in $E_{a,b}^{c,d}$. In other words, we have that $x - y \in E_a^{c,d} = E_a^c \subseteq G_a^c$ and thus x and y are also equivalent in $G_{a,b}^{c,d}$. Now to show that the map is injective, take $x, y \in E_b^{c,d}$ but equivalent as classes in $G_{a,b}^{c,d}$. Since $x - y \in E_b^{c,d}$ by the fact that $E_b^{c,d}$ is a vector space we have that $x - y \in E_b^{c,d} \cap G_{a,b}^{c,d} = E_a^c$. Thus they are equivalent in $E_{a,b}^{c,d}$ and indeed our map is injective, proving 3.6.5.

Proving $F^{c+\varepsilon,d-\varepsilon}_{a+\varepsilon,b-\varepsilon} \subseteq E^{c,d}_{a,b}$ For the last step of the proof, we show that:

$$\dim F^{c+\varepsilon,d-\varepsilon}_{a+\varepsilon,b-\varepsilon} \leq \dim E^{c,d}_{a,b} \tag{3.6.6}$$

by constructing a surjection from $E^{c,d}_{a,b}$ into $F^{c+\varepsilon,d-\varepsilon}_{a+\varepsilon,b-\varepsilon}$.

Recall that $E^{c,d}_{a,b} = \ker u_4 / \ker u_1$ and $F^{c+\varepsilon,d-\varepsilon}_{a+\varepsilon,b-\varepsilon} = \ker u_3 / \ker u_2$. We define a map from $E^{c,d}_{a,b}$ to $F^{c+\varepsilon,d-\varepsilon}_{a+\varepsilon,b-\varepsilon}$ by mapping a class x to $s_3(x)$. To show this map is well-defined, let x, y be equivalent classes in $E^{c,d}_{a,b}$, that is: $x - y \in \ker u_1$. We need to show that $s_3(x-y) = s_3(x) - s_3(y) \in \ker u_2$. But this is clear from commutativity of the diagram: $u_2 \circ s_2(\ker u_1) = u_3 \circ s_3(E^c_b) = u_3(\ker u_3) = 0$. So we map equivalent classes to equivalent classes and thus the map is well-defined.

Surjectivity of this map follows from the way we defined E^c_b, namely $s_3(E^c_b) = s_3(\ker u_1) = \ker u_3$. Thus the 3.6.6 holds, proving the theorem. □

As stated at the beginning of this section, we first prove a particular case of the Bottleneck Stability where we assume a certain restriction on the distance between f and g. To be more precise, we require f to be *very close* to g.

Definition 3.6.5. Let $\delta_f := \min\{\, \|p-q\|_\infty \mid p \in D(f) - \Delta, q \in D(f), p \neq q \,\}$ and let K be a simplicial complex. A function $g : K \to \mathbb{R}$ is said to be **very close to** $f : K \to \mathbb{R}$ when $\|f - g\|_\infty < \delta_f/2$.

Lemma 3.6.6 (Easy Bijection Lemma [1])**.** Let $f, g : K \to \mathbb{R}$ be monotonic functions on a simplicial complex K such that g is very close to f. Then:

$$W_\infty(D(f), D(g)) \leq \|f - g\|_\infty$$

Proof. We will prove the lemma by constructing a bijection from $D(g)$ to $D(f)$, such that the distance between its image and $D(f)$ is less than ε. Let p be a point in $D(f) - \Delta$. We define $R(p, \varepsilon)$ to be the square centered around p with side lengths ε. Let μ be the multiplicity of p in $D(f)$. Now by definition of $\varepsilon = \|f - g\|_\infty$, we have that $p \in (D(g) \cap R(p, \varepsilon))$ so by the box lemma:

$$\mu \leq \#(D(g) \cap R(p,\varepsilon)) \leq \#(D(f) \cap R(p, 2\varepsilon)) = \mu$$

The last equality follows from the assumption that g is very close to f. Indeed, this means that $2\varepsilon \leq \delta_f$ so that p is the only element in $D(f) \cap R(p, 2\varepsilon)$. Thus we can map all points in $D(g) \cap R(p, \varepsilon)$ to p, since its multiplicity now equals μ.

Doing this for all points in $D(f) - \Delta$, the leftover points P of $D(g)$ that are not yet mapped to $D(f)$ are further than ε away from $D(f) - \Delta$. We claim now that the distance of these points to the diagonal is less than or equal to ε. Let $(x, y) \in D(g)$. By the box lemma **??** and for δ small enough:

$$\begin{aligned} 1 &\leq \#(D(g) \cap [x-\delta, x+\delta] \times [y-\delta, y+\delta] && (3.6.7)\\ &\leq \#(D(f) \cap [x-\varepsilon, x+\varepsilon] \times [y-\varepsilon, y+\varepsilon] && (3.6.8) \end{aligned}$$

So for each point in $D(g)$, there is at least one point in $D(f)$ that is ε-close to it. Since each point in P is further than ε from $D(f) - \Delta$, it is closer than ε to Δ. Thus we finish our bijection by mapping each point in P to the closest point in Δ, keeping in mind that each point in Δ has infinite multiplicity. By construction, each point in $D(g)$ is mapped to a point in $D(f)$ that is not further than ε away, proving the lemma. □

We now prove the main theorem of the section. This is the general case where we drop the condition that g is very close to f.

Proof of theorem 3.6.2 Let $h_t := tf + (1-t)g$ be a linear interpolation between f and g for $t \in [0,1]$. Let σ be a face of τ in K. Then by monotonicity of f and g:

$$h_t(\sigma) = tf(\sigma) + (1-t)g(\sigma) \leq tf(\tau) + (1-t)g(\tau) = h_t(\tau)$$

So h_t is also monotonic for all t.

Let $c := \|f-g\|_\infty$ and $\delta(\lambda) := \delta_{h_\lambda}$. Then the sets $J_\lambda := (\lambda - \delta(\lambda)/4c, \lambda + \delta(\lambda)/4c)$ with $\lambda \in [0,1]$ form a cover of $[0,1]$. By compactness of $[0,1]$ there is a minimal finite subcover C of $[0,1]$. Let $\lambda_1 < ... < \lambda_n$ be the midpoints of the intervals J_{λ_i} in C. Since C is minimal $J_{\lambda_i} \cap J_{\lambda_{i+1}} \neq \emptyset$ so that by construction of J_λ we have:

$$\begin{aligned} \lambda_{i+1} - \lambda_i &\leq \delta(\lambda_{i+1}) + \delta(\lambda_i)/4c && (3.6.9)\\ &\leq \max\{\delta(\lambda_{i+1}), \delta(\lambda_i)\}/2c && (3.6.10) \end{aligned}$$

For $1 \leq i \leq n-1$. By minimality of C, we have $0 \in J_{\lambda_1}$ and $1 \in J_{\lambda_n}$ so that defining $\lambda_0 := 0$ and $\lambda_{n+1} := 1$ we have that the inequality also holds for $0 \leq i \leq n$.

To apply 3.6.6 to each pair $h_{\lambda_i}, h_{\lambda_{i+1}}$, we need that h_{λ_i} is very close to $h_{\lambda_{i+1}}$ or vice versa. Indeed this is the case. By definition of h_t and using inequality 3.6.9:

$$\begin{aligned} \|h_{\lambda_{i+1}} - h_{\lambda_i}\|_\infty &= (\lambda_{i+1} - \lambda_i)\|f-g\|_\infty && (3.6.11)\\ &\leq \max\{\delta(\lambda_{i+1}), \delta(\lambda_i)\}\frac{\|f-g\|_\infty}{2c} && (3.6.12)\\ &= \max\{\delta(\lambda_{i+1}), \delta(\lambda_i)\}/2 && (3.6.13) \end{aligned}$$

So $W_\infty(D(h_{\lambda_{i+1}}), D(h_{\lambda_i})) \leq \|h_{\lambda_{i+1}} - h_{\lambda_i}\|_\infty$.

Now using 3.6.6 and the triangle inequality:

$$\begin{aligned} W_\infty(D(f), D(g)) &\leq \sum_{i=0}^{n} W_\infty(D(h_{\lambda_{i+1}}), D(h_{\lambda_i})) && (3.6.14)\\ &\leq \sum_{i=0}^{n} \|h_{\lambda_{i+1}} - h_{\lambda_i}\|_\infty && (3.6.15)\\ &= \sum_{i=0}^{n}(\lambda_{i+1} - \lambda_i)\|f-g\|_\infty = \|f-g\|_\infty && (3.6.16) \end{aligned}$$

$\square$

Bibliography

[1] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. “Stability of persistence diagrams”. In: *Discrete & Computational Geometry* 37.1 (2007), pp. 103–120.

[2] Herbert Edelsbrunner and John L. Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.

[3] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. “Topological persistence and simplification”. In: *Proceedings 41st annual symposium on foundations of computer science*. IEEE. 2000, pp. 454–463.

[4] Davide L Ferrario and Renzo A Piccinini. *Simplicial structures in topology*. Springer Science & Business Media, 2010.

[5] Marian Gidea and Yuri A. Katz. “Topological Data Analysis of Financial Time Series: Landscapes of Crashes”. In: *SSRN Electronic Journal* (2017). DOI: `10.2139/ssrn.2931836`.

[6] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2019.

[7] Kathryn Heal. *Variations on the Nerve Theorem*. Mar. 2017. URL: `https://kathrynhealcom.files.wordpress.com/2017/03/final.pdf`.

[8] Leonid Polterovich et al. *Topological Persistence in Geometry and Analysis*. American Mathematical Society, 2020.

[9] André Weil. “Sur les théorèmes de de Rham.” In: *Commentarii mathematici Helvetici* 26 (1952), pp. 119–145. URL: `http://eudml.org/doc/139040`.